


```

? TITLE OF INVENTION: GLYCOSAMINOGLYCAN SPECIFIC
? TITLE OF INVENTION: SULFOTRANSFERASES
? NUMBER OF SEQUENCES: 19
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Fish & Richardson
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: Massachusetts
? COUNTRY: U.S.A.
? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
? COMPILER: IBM PS/2 Model 502 or 55SX
? OPERATING SYSTEM: MS-DOS (Version 5.0)
? SOFTWARE: WordPerfect (Version 5.1)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PC1/US93/05412
? FILING DATE: 19930607
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/899,432
? FILING DATE: 16 June 1992
? ATTORNEY/AGENT INFORMATION:
? NAME: CLARK, PAUL I.
? REGISTRATION NUMBER: 39,163
? REFERENCE/DOCKET NUMBER: 04020/015002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 542-5070
? TELEFAX: (617) 542-8906
? TELEX: 200154
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4052
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? PCT-US93-05412-1

alignment_scores:
    Quality: 290.00      length: 275
    Ratio: 1.790        Gaps: 11
    Percent similarity: 58.909   Percent identity: 29.091

alignment_block:
US-09-557-262-2_copy_53_311 x PCT-US93-05412-1
Align seq 1/1 to: PCT-US93-05412-1 from: 1 to: 4052
5 GluLeuProGlnThrIlelleGlyValArgLysGlyTyrThrAspAl 21
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
2253 CGCTTCACAAAGCGTGATCATCATTCGCCGGCCACAAAACAGCACCAACG 2302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
21 aLeuLeuGluMetLeuSerLeuHisProAspValAlaAla..... 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2303 GTTATTACTTTGTGGATATACAGGAGAGATTAAGAGATTAAGAGATA 2352
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 .. AlaGluAsnGluValHisPhePheAspTrpGluCisHisTyr 48
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2353 GGTTCAGAACATTTTACACATATTAAGTTTTITAAI...GGTACACATAT 2399
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 SerGlnGlyLeuGlyTyrTrpLeuThr.....GlnMetProPheSerSe 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2400 CAAAAAGGCATCGACGTGATACGAATTCCTTCCTATTTCCTTCACACAC 2449
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
63 rProHisGlnLeuThrValGluLysThrProAlaTyrPheThrSerProL 80
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2450 CACTCTGTATTTCTACTTTGAAAAAAGGTGGCAACTATTTTATTTCAGAG 2499
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
80 ysValProGluArgIleHisSerMetAsnProPhrIleArgLeuLeuLeu 96
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2500 TGAATATATATATATATATATATATATATATATATATATATATATATAT 2549
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: /cqn2.6/prodata/2/ins/pct/us93_05412_1
seq_documentation_block:
? Sequence 1, Application PC/TUS9305412
? GENERAL INFORMATION:
? APPLICANT: Hirschberg, Carlos B.
? APPLICANT: Orellana, Ariel
? APPLICANT: Hashimoto, Yasuhiro
? APPLICANT: Swiedler, Stuart J.
? APPLICANT: Wei, Zheng
? APPLICANT: Ishihara, Masayuki

```



```

61 eSerSerProHisGlnLeuThrValGluLysThrProAlaTyrPheThrS 78
157 ACCAACAGAGTACTTTGCTTCAAGCAATGCAACCAATTCAC 198
78 eProLysValProGluArgIleHisSerMetAsnProIleIleArgLeu 94
199 NAGAAATTCAGATTCTATATATATATATATATATATATATATAT 246
95 LeuLeuIleLeuArgAspProSerGluArgValLeuSerAspTyrThrG 111
247 CACAAATTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
111 nValLeuTyrAsnHisLeuGlnLysHisLysProTyrProProIleGlu 128
270 TATGTTTGGGACAC 310
128 sLeuLeuMetArgAspGlyArgLeuAsnLeuAspTyrIleValAlaLeu 144
311 CTATACCAAAATTTAAAGAGGTTTATTTTATTTTATTTTATTTT 354
145 ArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPhePheProLe 161
355 CAGAAGGCTGATCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 387
161 uGlyHisIleHisIleValAspGlyAspArg 175
388 ATTTCAAGGAGAGATGACAAAGAGGTCGCAAAAGG 424
175 sProPheProGluIleGlnLysValGluArgPheLeuLysLeuSerPro 191
425 ATTCTGAGAGAGATTAAGGATTTGAGAGTTTCTCTCTCTCTCTCTCT 474
192 GlnIleAsnAlaSerAsnPheTyrPheAsnLysThrLys 206
475 TATTTGGGGTGACACATTTGGCTTCTGATCTCTCTCTCTCTCTCT 524
206 eTyrCys 208
525 CTATTGC 531

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq.us-08-441-139-17

seq_documentation_block:

; Sequence 17, Application US/08441119

; Patent No. 5773245

; GENERAL INFORMATION:

; APPLICANT: Wittrop, Dr. Karl D.

; APPLICANT: Robinson, Ann S

; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF

; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS

; NUMBER OF SOURCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESS: SCULLY, SCOTT, MURPHY & PRESSER

; STREET: 400 Garden City Plaza

; CITY: Garden City

; STATE: NY

; COUNTRY: USA

; ZIP: 11530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US 08/084,947

; FILING DATE: 06-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Digiglio, Frank S.

; REGISTRATION NUMBER: 31,346

```

; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELETYPE: 230 901 SANS OR
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-441-139-17

```

alignment_scores:

```

Quality: 84.50 length: 166
Ratio: 1.006 Gaps: 8
Percent Similarity: 50.402 Percent identity: 25.301

```

alignment_block:

US-09-557-262-2_copy_53_311 x US-08-441-139-17

Align seq 1/1 to: US-08-441-139-17 from: 1 to: 2403

```

105 ValLeuSerAspTyrThrGln.....ValLeuTy 114
1186 GTTTCGGCCCAATACGTCGAAAGCGGTTTCCTTTGGTTACTTGTCTA 1235
114 rAsnHisLeuLysHisLysProTyrProProIleGluAspLeuM 131
1236 CAATGAGAGAGAAATTTGAAAGATTAACAGATCTCTTTTACAGAGTGG 1285
131 cTatAspTyrArg.....LeuAsnLeuAspTyrLysAla 142
1286 CCAAAAAGACAGAGGCTTAATGAACCTTTGTTAGCATGATGACGAGAAA 1335
143 LeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPhe 159
1336 TTCCGTAGAG.....CATGGGGGCAACTTCAACATGAGGAGCAACAT 1376
159 eProLeuGlyHisIleHisIleValAspGlyAspArgLeuIleArgAsp 176
1377 CCTCTTATTTCGAAATGAGATGATGATGAGAA.....TTGAAATAG 1420
176 rPheProGluIleGlnLysValGluArgPheLeuLysLeuSerThr 132
1421 GTTTCATTAATTTCTCTTAA.....GAGAGGTTTGAATTAATGAGACA 1467
193 Ile...AsnAlaSerAsnPheTyrPheAsnLysThrLysGlyPheTyr 208
1468 ATCGGTCTTGGAGTCCCAAGGCTATTCAATCTTTGGTTAAGGACTTCTG 1515
208 sLeuArgAspSerGlyLysArgArgCysLeuHisGluSerLysGlyArg 225
1516 ..... 1525
225 tHisProGlnValAspProLysLeuLeuAspLysLeuHisGluTyrPhe 241
1526 CATTGAGAAATTTG 1557
242 HisGluProAsnLysLysPhePheLysLeuValIleArgThrPheAsp 257
1558 CAGAACTAAGATTTCTGTTTCTCAATGCTGTTGTAAGCAACAGAC 1605

```

seq_name: /cgn2_6/ptodata/2/ina/6b_COMB.seq.us-09-056-556-183

seq_documentation_block:

; Sequence 183, Application US/09056556

; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skelky, Yasir A.W.


```

141 lysAlaIleuAsnArgSerLeuTyrHisAlaHisMetIleuAsnTrpIleuAr 157
      :::::  :::::
1481 ATGGGATATCTCCGAGCTGAG      AGGTTAGAGAGTGGATGAG 1441
      :::::  :::::
157 qPhePheProLeuGlyHisIleHisIleValAspGlyAspArgLeuIleA 174
      |  :::::  :::::
1440 GCGCGCGCGCGAGCTGGAGGAGAACAGCGCTGGAGAGCTCACAGCTGATT 1391
      :::::  :::::
174 rgAspProPhePro 178
      :::::  :::::
1390 TCTCTCCATACCCC 1377

```

```

seq_name: /cgr2_6/ptodata/2/ina/5A_COMB.seq US-08-727-034-2

```

```

seq_documentation_block:
; Sequence 2: Application US/08727034
; Patent No. 5663872
; GENERAL INFORMATION:
; APPLICANT: SATTO, YASHUSHI
; APPLICANT: IWASAKI, AKIO
; APPLICANT: ARAI, KOICHI
; APPLICANT: YAMAZAKI, HIROYUKI
; TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
; TITLE OF INVENTION: THE GENE CODING THEREFOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GELON, SPIVAK, McLELLAND, MATER & NEUSTADT.
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,034
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 261440/1995
; FILING DATE: 09-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 102451/1996
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OHLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 80-0/9-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6961 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 178..6961
; OTHER INFORMATION: /note "Identification Method: S"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 178..261
; OTHER INFORMATION: /note "Identification Method: S"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 262..6816

```

```

; OTHER INFORMATION: /function- "Nucleotides 262-6816
; OTHER INFORMATION: encode the mature peptide"
; OTHER INFORMATION: /note- "Identification Method: S"
US-08-727-034-2

alignment_scores:
      Quality: 79.50      Length: 172
      Ratio: 0.970      Gaps: 7
Percent Similarity: 47.674      Percent Identity: 26.163

alignment_block:
US 09 557 262 2 copy 53 311 x US-08 727 034 2 rev
      10 IleIleIleGlyValArgLysGlyGlyThrArgAlaLeuLeuMetLe 26
      :::::  :::::
1992 CTCCTTGTGGAGGCAAGATGGTGAAGACCGTGTCTCTCTC..... 1951
      26 uSerLeuHisProAspValAlaAlaAlaGluAsnGluValHisPhePheA 43
      :::::  :::::
1950 GCGGCTTCCTGTTGAGAGAGCCCATACACAAACAG 1917
      43 sPTTP.....GluGluHisTyrSerGlnGlyLeuGlyTyrTyrLeuThr 57
      :::::  :::::
1916 GATTTCTAGAGAAAGAGAGGCTTTTCTGAGCTTCTGCTTCTGTTACT 1867
      58 GluMetProPheSerSerProHisGluLeuThrValGluLysThrProAl 74
      :::::  :::::
1866 GTACTTCAGTTCTGCTTCTTCCATGCGCTGGGCAAGGGGATGATGATGC 1817
      74 dTyPheThrSerProLysValProGluAlaGluIleHisSerMetAsnProT 91
      :::::  :::::
1816 GCGATGTCCTCCCATGTATA.....G.....G 1794
      91 hrIleArgLeuLeuLeuLeuArgAspProSerGluArgValIleuSer 107
      :::::  :::::
1793 TAGTGAGTCCAGGAGATGGCTCTGCGCACTGTGCTGACACTGTGTAGA 1744
      108 Asp TyrThrGluValLeuTyrAsnHisLeuGlnGlyHisLeuLysProTyrP 124
      :::::  :::::
1743 CATGACACACGTTTGTCTTG.....GAGCCCAAGTCTTCTTCCACATG 1703
      124 rProIleGluAspLeuMetArgAspGlyArgLeuAsnLeuAspTyr 140
      :::::  :::::
1762 AGAGAGTGGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
      141 LysAlaLeuAsnArgSerLeuTyrHisAlaHisMetIleuAsnTrpIleuAr 157
      :::::  :::::
1658 ATGGGATATCTCCGAGCTGAG      AGGTTAGAGAGTGGATGAG 1618
      157 qPhePheProLeuGlyHisIleHisIleValAspGlyAspArgLeuIleA 174
      |  :::::  :::::
1617 GCGCGCGCGCGAGCTGGAGGAGAACAGCGCTGGAGAGCTCACAGCTGATT 1568
      :::::  :::::
174 rgAspProPhePro 178
      :::::  :::::
1567 TCTCTCCATACCCC 1554

```

```

seq_name: /cgr2_6/ptodata/2/ina/5A_COMB.seq US-08-727-034-2

```

```

seq_documentation_block:
; Sequence 1: Application US/08363300
; Patent No. 5700927
; GENERAL INFORMATION:
; APPLICANT: ZON, Leonard and Richardson, Paul
; TITLE OF INVENTION: Tbet Gene and Uses Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston

```

```

: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: ProDOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/363,300
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 04590/009001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4039 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 403..3829
: US-08-363-300-1

```

```

alignment_scores:
  Quality: 79.00 Length: 108
  Ratio: 1.837 Gaps: 5
  Percent Similarity: 39.815 Percent Identity: 25.926

```

```

alignment_block:
  US-09-557-262-2_COPY_53_311 x US-08-363-300-1

```

```

Align seq 1/1 to: US-08-363-300-1 from: 1 to: 4039

```

```

48 TyrSerGlnGlyLeu ..... GlyTyrTyrLeuThrG1 58
.....
2901 CATCTCAACAGCCATATCTGCTCTGACACACACAGCTGGATAC1022CAAG 2950
58 nMetPro..... PheSerSerProHisGln ..... 66
: .....
2951 GTCACGATTTCGAGAGGATTTCTTTCTTCAATCATGATGATGAGAGAG 4000
67 .....
3001 GCCTTCACAGATGCTCAAGTTCCTGCACTTTCACATGAGGCGGCGGCAACAA 4050
69 ValGluLysThrProAlaTyrPheThrSerProLysValProGluArg11 85
.....
3051 GTATCGGCCACAGATATGATTTTTCAGATCATACATGTA ..... 3099
85 eHisSerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProS 102
: .....
3090 .....
102 er.GluArgValIleSerMetAspTyrThrGlnValValTyrAsnHisLeuG1 118
: .....
3094 CTCTACACGCTCTCTTACCATATTCACATACACACATCTTCTTCAACACATCTTCA 3143
118 nLysHisLysProTyrProPro 125
: .....
3144 AGACCACGACACGCGGCCCCCT 3165

```

```

seq_name: /cgn2_6/p2data/2/ins/557_557_COPY_53_311 x US-08-363-300-1

```

```

seq_documentation_block:

```

```

: Sequence 1: Application US/08/249,687C
: Patent No. 5942412
: GENERAL INFORMATION:
: APPLICANT: PRAGER, DIANE
: APPLICANT: MELMED, SHLOMO
: TITLE OF INVENTION: POLYNUCLEIC ACID ENCODING
: TITLE OF INVENTION: VARIANT INSULIN LIKE GROWTH FACTOR I RECEPTOR BETA
: TITLE OF INVENTION: SUBUNIT & RECEPTOR
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pretty, Schroeder & Poplawski
: STREET: 444 South Flower St., 19th Floor
: CITY: Los Angeles
: STATE: CA
: COUNTRY: USA
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/249,687C
: FILING DATE: 26-MAY-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/044,540
: FILING DATE: 06-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Basile, Lega
: REGISTRATION NUMBER: P-44,026
: REFERENCE/DOCKET NUMBER: 007 42449
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 213-622-7700
: TELEFAX: 213-489-4210
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4975 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: both
: MOLECULE TYPE: cDNA
: US-08-249-687C-1

```

```

alignment_scores:
  Quality: 79.00 Length: 146
  Ratio: 1.197 Gaps: 9
  Percent Similarity: 47.143 Percent Identity: 26.429

```

```

alignment_block:
  US-08-249-687C-2_COPY_53_311 x US-08-249-687C-1/rev

```

```

Align seq 1/1 to reverse of: US-08-249-687C-1 from: 1 to: 4975

```

```

54 TrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuPheValG1 70
: .....
3116 TGGTTCAGGTCATCTT.....TCACACACACCTTGGCACTTCT 4076
70 nLysThrProAlaTyrPheThrSerProLysValPro.....Glu 83
: .....
3075 TCATACATCATCCCAACACACCTTGGCACTTGGCACTTGGCACTTGGCA 4926
94 ArgIleHisSerMetAspProThrIleArgLeuLeuLeuLeuArgAs 109
: .....
3025 CTCTACACGCTCTCTTACCATATTCACATACACACATCTTCTTCAACAC 2984
100 pProSerGluArgValIleSerMetAspTyrThrGlnValIleuLysHis 117
: .....
2983 AGACACGCTCTCTTACCATATTCACATACACACATCTTCTTCAACAC 2950
117 euGlnLysHisLysProTyrProProLleGluAspLeuLeuMetArgAsp 133

```

```

2949 TACAGCACATCCCGCCAGCCGCTGTTATTTCTCTTCTTCTATGGAAGAC 2900
134 GlyArgLeuAsn.LeuAspTyrLysAlaLeuAsnArgSerLeuTyrHis 150
2899 GTACAGCATAATCACCACCAACCTTCCCAAGATCAACACGACACAGCGGGCA 2850
150 LeHisMetLeuAsnTrpLeuArgPhe...PheProLeu..... 161
2849 GAGCGATGATCAGATGGATGAAGTTTTCATATCTGTGTTTGGGCTGGACA 2800
162 .....GlyHisIleHisIleValAspClyAspArgLeuIleArg 174
2799 TAGAAGAACACAGCATCTCTCCAC..... 2776
174 AspProPheProGlu 179
2775 .GAGCCATTCCACAG 2761
```


XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX Homo sapiens.
 OS
 PN W0200157276 A2.
 PN 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-050066A
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0634466.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000US-0024263.
 XX
 PA (MOLE) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank D9;
 XX
 XX WPI; 2801 488400/53
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 16679; 658pp + Sequence listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 SQ Sequence 924 BP; 169 A; 252 C; 310 G; 104 T; 0 other;

alignment_scores:

Quality: 1277.00 Length: 253
 Ratio: 5.129 Gaps: 0
 Percent Similarity: 98.419 Percent Identity: 92.095

alignment_block:

US-09-557-262-2_copy_53_311 x AAK42122/rev ..

Align seq 1/1 to reverse of: AAK42122 from: 1 to: 924

1 GlySerThrGlnGlnLeuProGlnThrIleIleIleGlyValArgLysG1 17
 761 GGCCTCCGCCAGCAGTTGCGGCACACCAATCAATCGGCGTGGCAAGGG 712
 17 yClyThrArgAlaIleuLeuGluMetCysSerLeuHisProAspValAla 34
 711 GGGTAAAGAGGAAATGATGAGATGAGATGAGATGAGATGAGATGAG 662
 34 IaAlaGluAsnGluValHisPhePheAspTrpGluGluHisTyrSerGln 50
 661 CCGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
 51 GlyLeuGlyTrpTyrLeuThrIleMetProPheSerSerProHisGlnLe 67
 611 GGGTGGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 562
 67 nThrValGluGlyThrProAlaIleuPheThrSerCysGlySerValProGln 84
 561 CAAAGTGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
 84 rGlnHisSerMetAsnProThrIleArgGluLeuLeuLeuArgAsp 100

511 GAGTCTACAGCATGAACCGCTCCATCCGGCTGCTCTCTCTCTCTCTCT 462
 101 ProSerGluArgValLeuSerAspTyrThrGluValLeuTyrAsnIleLe 117
 461 CCGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
 117 uGluGlyHisLysProTyrProGlnIleGluAspLeuLeuMetArgAspG 144
 411 GGAAGAAATGAAGGATGAGATGAGATGAGATGAGATGAGATGAGATG 362
 134 yArgLeuAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAla 150
 361 GCAAGCTTCAATGTGGACTACAAAGGCTTCAAGGCTTCAAGGCTTCA 412
 151 HisMetLeuAsnTrpLeuArgPhePhePheLeuGlyHisIleIleVal 167
 311 CAAATGAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 262
 167 AspGlyAspArgLeuIleArgAspProPheProGluIleGlnIysValG 184
 261 GAAAG 212
 184 LuArgPheLeuLysLeuSerProGlnIleAsnAlaSerAspPheTyrPhe 200
 211 AGAGCTTCTTAAAGGCTTCTTAAAGGCTTCTTAAAGGCTTCTTAAAG 162
 201 AsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgG 217
 161 AATAAATAAAGGCTTCTTAAAGGCTTCTTAAAGGCTTCTTAAAGG 112
 217 LeuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
 111 GTTAAATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 62
 234 euAspLysLeuHisGlyTyrPheHisGluProAsnLysLysPhePheLys 250
 61 TCATTAACATCCAGCAATATTTTCATGAGCCAAATAAAGAAATCTCT 12
 251 LeuValGly 253
 11 CTGTTGGC 3
 seq_documentation_block:
 ID AAI48186 standard; DNA: 924 BP.
 XX
 AC AAI48186;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #16872 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human, placenta, antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-050066A
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0634466.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000US-0024263.
 XX
 PA (MOLE) MOLECULAR DYNAMICS INC.

Claim 5; Page 152-153; 158pp; English.

PS The present sequence is human drug metabolising enzyme (DME)-5 cDNA.
 XX DME polypeptide, polynucleotide and modulators are useful for
 CC diagnosis, treatment and prevention of autoimmune/inflammatory,
 CC cell proliferative, developmental, endocrine, eye, metabolic,
 CC and gastrointestinal disorders, including liver disorders.
 CC The autoimmune/inflammatory disorders treatable include
 CC AIDS, adult respiratory distress syndrome, Addison's disease,
 CC allergic anaphylaxis, asthma, atherosclerosis, osteoporosis, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, Crohn's disease, atopic
 CC dermatitis, diabetic mellitus, Graves' disease, glomerulonephritis,
 CC rheumatoid arthritis, scleroderma, systemic lupus erythematosus,
 CC systemic sclerosis, ulcerative colitis, haemodialysis and uveitis.
 CC viral, bacterial, fungal, parasitic, protozoan, helminthic infections
 CC and trauma, and cell proliferative disorders such as cancer, actinic
 CC keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,
 CC hepatitis and psoriasis. Developmental disorders include anaemia, renal
 CC tubular acidosis, epilepsy, hypothyroidism and cataract, and endocrine
 CC disorders include disorders of hypothalamus and pituitary, disorders
 CC associated with hypopituitarism, including sarcoidosis, diabetes
 CC insipidus, hypogonadism, disorders associated with hypothyroidism
 CC including goitre, acute thyroiditis, Graves' disease, disorders
 CC associated with hyperparathyroidism, pancreatic disorders such as type I
 CC or type II diabetes mellitus, disorders associated with adrenals such as
 CC hyperplasia, Cushing's disease, endometriosis, infertility.
 CC hypergonadal disorders, and gynaecomastia. Eye disorders include
 CC conjunctivitis, keratitis, glaucoma and macular degeneration, and
 CC metabolic disorders include diabetes, cystic fibrosis, goitre,
 CC hyperbatespermatemia, hypothyroidism, hyperlipidaemia, lysosomal storage
 CC diseases, obesity, phenylketonuria and hypocalcaemia. Also the molecules
 CC are useful for treating gastrointestinal disorders such as dysphagia,
 CC gastritis, peptic ulcer, cholelithiasis, cirrhosis, hepatitis,
 CC hyperbilirubinaemia, constipation, diarrhoea, jaundice, Wilson's disease,
 CC thrombosis and hepatic tumours. The DME polypeptide is also useful for
 CC screening its agonist or antagonist.

SQ Sequence 2115 bp: 651 A: 408 C: 438 G: 618 T: 609 other:

alignment_scores:
 Quality: 755.90 Length: 258
 Ratio: 3.564 Gaps: 2
 Percent Similarity: 82.171 Percent Identity: 53.488

alignment_block:

US-09-557-262-2_copy_53_311 x AAD24010 ..

Align seq 1/1 to: AAD24010 from: 1 to: 2115

4 GlnGlnIouProGlnIhrIleIleIleIleGlyValArgIysGlyGlyThrAr 30
 438 CAGCAGCCTCCGAGGAGGATATCATTCGGGTCAGGAAAGAGGAGGAGCAAC 487
 20 qAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaAlaGluA 37
 488 GATCTGTGTGAAATGCTGAACTTACATTCGGGAGTAGTCAAAAGGCTCTC 537
 37 snGlnValHisPheAspTrpGluHisIstyrSerGlnGlyIleuGly 53
 538 AAGAAATCCACTTTTGTATATGATGAGAAATTAIGTAAAGGAGGATGAG 587
 54 TrpIyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValGI 70
 588 TGGATATAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 637
 70 dLysThrProAlaTyrPheThrSerProLysValPleGluAlaIleHis 87
 638 AAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 687
 87 crMetAsnProThrIleArgIleuLeuLeuLeuArgAspProSerGlu 103
 688 AATATGAA-TTATGATGAA-TTATGATGATGATGATGATGATGATGATGAT 737

104 ArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnIstysHI 120
 738 AGAGCTATTTCTGATTAATCTCAGGTCAGGAGGAGGAGGAGGAGGAGGAG 787
 120 stypProTyrProProIleGluAspLeuLeuMetArgAspGly.....A 145
 788 CAAATTTATTCAAGATTTTGAAGAGCTGGGCAATAGACGATTAATACAGG 837
 135 rgLeuAsnLeuAspTyrIstysAlaLeuAsnArgSerIleuTyrHisAlaHis 151
 838 AAGTGAACACAAAATAAAGAGAGTAAGACAGCAATACACAAACAT 887
 152 MetLeuAsnTrpLeuAlaPhePheProLeuGlyIleIleIleIleValAs 168
 888 CTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 937
 168 pGIYAspArgLeuIleArgAspProPheProGluIleGlnLysValGluA 185
 938 TGAATATGAGGATATATATATATATATATATATATATATATATATAT 987
 185 rgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheIstysPheAsn 201
 988 AGTCTCTAATTCCTGCTCCAGGATATATATATATATATATATATATAT 1037
 202 LysThrLysLysPheTyrCysLeuArg...AspSerGlyLysAspArgGly 217
 1038 CTTACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1087
 217 sLeuHisGluSerLysGlyArgAlaHisProGluValAspProLysLeuL 244
 1088 CTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137
 234 euAspLysLeuHisGlnIstysPheHisGlnIstysPheHisGlnIstys 250
 1138 TTACTAATTCGCAATTCCTTCATCTCTTTTAAATCAAAAATTTTACGAG 1187
 251 LeuValGlyAlaThrPheAspTrp 258
 1188 ATCCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1211

seq_name: /SILS5/seqdata/geneseq/geneseq_emb1/RA2001A.DAT:ABA44908

seq_documentation_block:

ID ABA44908 standard; DNA; 452 bp.

AC ABA44908;

XX 01-FEB-2002 (first entry)

XX Human breast cell single exon nucleic acid probe #3603.

XX Human: microarray; single exon probe; gene expression; breast; disease; cancer; ss.

XX Homo sapiens.

XX W0200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001W9-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632466.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000CB-0324263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-496933/54.
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 1: SEQ ID NO 3603; 427pp + sequence listing; English.
 XX
 CC The invention relates to a spatially addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and H1 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or producing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_COPY_54_411 x ABA44406/rev

Align seq 1/1 to reverse of: ABA444908 from: 1 to: 452

37 AsnGluValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuGI 53
 |||||
 451 ACGAGGTCACCTTTCTGACTGGGAGGAGATTACAGCCAGCGCTGGG 402
 53 YTrpTyrLeuThrGlnMetProPheSerSerProHisnLeuThrValG 70
 |||||
 401 CTGCTACCTTCACCCACCTCCCTCTCTGCGCCACACCACTCACACTGG 352
 70 LuLysThrProAlaTyrPheThrSerProLysValProGluArgIleHis 86
 |||||
 351 AGAAGAGCGCGCGCGTATTATATGTTGGGAAATGCTGTAAGATCTAT 402
 87 SerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAspProSerGI 103
 |||||
 301 AGATGAAAGCTATATGCGGTCTGTCTATCTATGTAAGATGATGGA 252
 103 uArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 120
 |||||
 251 CGCGCTCTTCTATCTGACTACAGCCCACTCTCTACAAACACATGAGAAC 202
 120 isLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeu 136
 |||||
 201 ACAAGGCTATAGGATGATATGATAGGAGATCTGATGATGATGATGAT 152
 137 AsnLeuAspTyrTyrAlaIleuAsnArgSerLeuTyrHisAlaHisMetLe 153
 |||||
 151 AATGHisA-TA-AAAGGCTTAAAGG-AGCTTAAAGCTTAAAGCTTAAAG 102
 153 uAsnTrpLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 176
 |||||
 101 GAACG 92

170 SPATGLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 186
 |||||
 51 ACGAGGTCACCTTTCTGACTGGGAGGAGATTACAGCCAGCGCTGGG 2

seq_name: /SIDS/qcdata/geneseq/geneseq_emb1/NA2001A.DAT.AHA55377

seq_documentation_block:

ID ABA55377 standard; DNA; 452 BP.

AC ABA55377;

XX 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #3682.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX W0200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001W-0500669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0612366.

XX 21-SEP-2000; 2000US-0234887.

XX 27-SEP-2000; 2000US-0235359.

XX 04-OCT-2000; 2000EB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 1: SEQ ID NO 3682; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:

Quality: 748.00 Length: 150

Ratio: 5.123 Gaps: 0

Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_COPY_53_311 x ABA55377/rev

Align seq 1/1 to reverse of: ABA55377 from: 1 to: 452

37 AsnGluValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuGI 53

|||||

451 ACGAGGTCACCTTTCTGACTGGGAGGAGATTACAGCCAGCGCTGGG 402

53 YTrpTyrLeuThrGlnMetProPheSerSerProHisnLeuThrValG 70

Align seq 1/1 to reverse of: AAK29079 from: 1 to: 452

```
37 AsnGlnValHisPhePheAspTrpGlnGlnHisTyrSerGlnGlyLeuGI 53
|||||
451 AAGGAGGTGCAATTGTCGATGAGAGAGATTAAGAGAGATTTGAG 402
|||||
53 YTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 70
|||||
401 CTGGTACCTTCACCCAGATGCTTTGTCGCTGCTACACCACTACACTGG 352
|||||
70 LuLysThrProAlaTyrPheThrSerProLysValProGluArqIleHis 86
|||||
351 ACAAGACGCGGAGGAAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
|||||
87 SerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProSerGI 103
|||||
301 AGCATGAAAGGCTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 252
|||||
103 uArqValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 120
|||||
251 GCGCGTCTATCTGACTACACCCAGAGCTGCTGCTGCTGCTGCTGCTG 202
|||||
120 IsLysProTyrProIleGluAspLeuLeuMetArgAspGlyArqLeu 136
|||||
201 ACAAAGGCTATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 152
|||||
137 AsnLeuAspIleLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetL 153
|||||
151 AATGTGAGATTAAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAG 102
|||||
153 uAsnTrpLeuArqPheProLeuGlyHisIleHisIleValAspGlyA 170
|||||
101 CAACCTGGCTGGCGCTTTTTCGGCCCTGGCCCACTCCACATTCCTGAC 52
|||||
170 spArqLeuIleArqAspProPheProGluIleGlnLysValGluArqPhe 186
|||||
51 AGCGGCTGATAGAGAGAGATTAAGAGAGATTAAGAGAGATTAAGAG 2
```

OM of us 09-557-262-4_copy_49_307 Loc. GenBank: * out_format: .pfs
 Date: Jul 17, 2002 5:56 AM
 About: Results were produced by the GenCore software, version 4.5.
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame2.p2n.model -DEV=xlp
 -U/Cp21/Uspt0.spool/US09557262-4_copy_49_307 -runal 1-073002-161204-22557/arg_query fasta_1_1558
 -DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -DOPEXT=0.000 -DOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cgi -LIST=45 -DOALIGN=200 -TUR_SCORE=pcr
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT .pfs
 -NORM=ext -HEADS/ZE=500 -MINFN=0 -MAXFN=2000000000
 -USR=US09557262-4_copy_49_307 -CPU=5 -CPU=3 -LONGLOC
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-557-262-4_copy_49_307

Query length: 259

Database: GenBank *

Database sequences: 1797656

Database length: 187333701

Search time (sec): 7102.980000

score_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_pr:AF019386	+	1415.00	2604.31	1.1e-136	1305	AF019386 Homo sapiens heparan
gb_pr:AF073991	-	1415.00	2556.39	5.0e-134	156838	AF073991 Homo sapiens BA
gb_pr:AF006236	+	1398.00	2542.37	2.8e-133	166398	AF006236 Homo sapiens chromo
gb_to:AF009133	+	1318.00	2422.02	1.9e-126	1657	BC009133 Mus musculus, heparan
gb_to:AF019385	+	1318.00	2421.85	1.9e-126	1655	AF019385 Mus musculus, heparan
gb_hlg:AC084416_3	+	1318.00	2480.20	3.2e-124	118000	AC084416_3 Contiguation (4 of 6) of AC0
gb_hlg:AC098713	+	1310.00	2473.29	7.9e-124	22128	AC098713 Mus musculus, heparan
gb_to:AF177430	+	1309.00	2411.02	6.2e-126	936	AF177430 Rattus norvegicus 3-O-
gb_pat:AX317978	+	772.50	1407.81	4.7e-70	2115	AX317978 Sequence 18 from patc
gb_to:AL355498	+	667.00	1326.14	6.2e-60	519	AL355498 Human DNA sequence
gb_om:CA388534	+	598.00	1084.87	4.6e-52	1968	AF105374 Homo sapiens heparan
gb_pr:AF105374	+	598.00	1084.87	4.6e-52	1968	AF105374 Homo sapiens heparan
gb_to:AF168992	+	586.00	1064.28	6.4e-51	1666	AF168992 Mus musculus D-glycos
gb_pr:AF105377	+	585.00	1060.44	1.1e-50	2032	AF105377 Homo sapiens heparan
gb_pr:AF105376	+	585.00	1058.40	1.4e-50	2546	AF105376 Homo sapiens heparan
gb_pr:AF023723	+	585.00	1057.22	1.6e-50	2808	AF023723 Homo sapiens cDNA F3
gb_pr:AF105378	+	555.00	1003.96	1.5e-47	2211	AF105378 Homo sapiens heparan
gb_hlg:AF017305	+	460.50	790.72	1.1e-45	94745	AF017305 Drosophila melanogas
gb_to:AF008337	+	456.50	786.42	2.2e-45	16660	AF008337 Drosophila melanog
gb_to:AF003511	+	460.50	778.39	5.4e-35	343807	AF003511 Drosophila melanog
gb_hlg:AF027102	+	447.00	760.44	5.4e-34	18672	AF027102 Homo sapiens chromo
gb_pr:AF009237	+	447.00	759.43	6.3e-34	188636	AF009237 Homo sapiens chromo
gb_hlg:AC0093359	+	440.00	744.88	4.0e-33	218502	AC0093359 Mus musculus clone
gb_hlg:AC008715	+	440.00	743.57	4.7e-33	245269	AC008715 Mus musculus chromo
gb_pr:AF094511	+	438.00	744.46	4.2e-33	186234	AF094511 Homo sapiens chromo
gb_hlg:AC021445	+	428.50	724.25	5.6e-32	203621	AC021445 Mus musculus clone
gb_to:CHLH5803	+	426.50	748.96	2.5e-33	13172	AF098987 Caenorhabditis eleg
gb_to:AF098987	+	426.50	741.53	6.1e-33	24810	AF098987 Caenorhabditis eleg
gb_pr:AF006640	+	425.00	726.88	4.0e-32	81579	AF006640 Homo sapiens 16p13.3
gb_pr:HS430A6	+	423.00	720.97	2.7e-32	41255	AF031723 Human DNA sequence
gb_hlg:AL603846	+	421.00	705.84	5.9e-31	320046	AF031723 Human DNA sequence
gb_pr:AC005363	+	420.00	718.43	1.2e-31	75108	AC005363 Homo sapiens chromo
gb_hlg:AC012180	+	420.00	706.40	5.4e-31	251997	AC012180 Homo sapiens chromo
gb_pr:AC005375	+	419.00	709.74	3.6e-31	149330	AC005375 Homo sapiens chromo
gb_pr:AC005374	+	419.00	708.43	4.2e-31	166687	AC005374 Homo sapiens chromo
gb_to:CELF52B10	+	365.50	624.13	2.1e-26	38024	U01990 Caenorhabditis elegans
gb_to:AB036429	+	351.50	622.45	2.7e-26	3359	AF036429 Homo sapiens NUS14 mR
gb_to:AF175689	+	347.00	610.14	1.4e-25	4947	AF175689 Drosophila melanogast
gb_to:AB036838	+	346.50	615.13	6.7e-26	2732	AF036838 Mus musculus NUS14 mR

seq_name: gb_pr:AF019386
 seq_documentation_block:
 LOCUS AF019386 1305 bp mRNA linear F01 14-NOV-1997
 DEFINITION Homo sapiens heparan sulfate 3-O-sulfotransferase-1 precursor
 (3OST1) mRNA, complete cds.
 ACCESSION AF019386
 VERSION AF019386.1 GI:2618972
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1305)
 Shworak N.W., Liu J., Fritze L.M.S., Schwartz J.J., Zhang L.,
 Loefer D. and Rosenberg P.D.
 TITLE Molecular cloning and expression of mouse and human cDNAs encoding
 heparan sulfate D-glucosaminyl 3-O-sulfotransferase
 J. Biol. Chem. 272 (44), 28009-28019 (1997)
 REFERENCE 2 (bases 1 to 1305)
 Shworak N.W., Liu J., Fritze L.M.S., Schwartz J.J., Zhang L.,
 Loefer D. and Rosenberg P.D.
 TITLE Direct Submission
 JOURNAL Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
 02139, USA
 FEATURES
 Location/Qualifiers
 1..1305
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Isolated from lambda Triplex Brain cDNA library
 from Clontech"
 1..1305
 /gene="3OST1"
 /name="3OST1"
 1..118
 /gene="3OST1"
 119..178
 /gene="3OST1"
 119..1042
 /gene="3OST1"
 /function="rate limiting enzyme for synthesis of
 anticoagulant heparan"
 /note="heparan sulfate sulfotransferase; 3-OST-1;
 interluminal Golgi resident protein (retention mechanism
 unknown)"
 /codon_start=1
 /product="heparan sulfate 3-O-sulfotransferase-1
 precursor"
 /protein_id="AA084488.1"
 /db_xref="GI:2618973"
 /translation="MAALGLGAVLVAGLVFVFAELGQVQLRRKAGTLQDDVROG
 VAPNSACGLPTIIIGVKGKFTALLENLSLPEVAAAEVHFVFWEEYHSLGW
 YLSQMPFPHQITVEKTPAYTPKPVVSMNPSILILLIKDPSVLSYTOV
 FYNMGKKRPVSLFPIVFGKLVAVKALNPSLYHVHMNMLRFFPLRHHIVDGD
 ELKQFELKQVFFKLSQVFNKSNPFPKNGVYVPSQKPLHFSKQAHVQ
 VPKPLINKLHFVHFHPKPKFELVGRFHW"
 179..1039
 /gene="3OST1"
 /product="heparan sulfate 3-O-sulfotransferase-1"
 263..268
 /gene="3OST1"
 /note="encodes potential N-linked site"
 279..1536
 /gene="3OST1"
 /note="encodes presumptive sulfotransferase catalytic
 domain"
 527..535
 misc_signal

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The RPO1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frengren, E., Tatem, M., Catasuso, J. J., and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-669R3; the clone sequenced to the right is AC006230. Actual start of this clone is at base position 1 of RP11-512I20; actual end is at base position 159838 of RP11-512I20.

repeat_region	11853..12152	/rpt_family="MaLR"
repeat_region	12209..12271	/rpt_family="L1"
repeat_region	12305..12390	/rpt_family="L2"
repeat_region	12605..12876	/rpt_family="L1"
repeat_region	12877..13185	/rpt_family="Alu"
misc_feature	13162..13185	/note="similar to Homo sapiens EST BF183502 (NM_011061813)"
repeat_region	13186..13221	/rpt_family="L1"
repeat_region	13223..13488	/rpt_family="L1"
repeat_region	13489..13750	/rpt_family="Alu"
repeat_region	13751..14291	/rpt_family="L1"
repeat_region	14332..14380	/rpt_family="MERL_type"
repeat_region	15690..15714	/rpt_family="MERL"
repeat_region	15873..16172	/rpt_family="(T)n"
repeat_region	16173..16300	/rpt_family="Alu"
repeat_region	16910..16942	/rpt_family="Alu"
repeat_region	17130..17499	/rpt_family="AT-rich"
repeat_region	17303..17379	/rpt_family="ERVL"
repeat_region	17553..17621	/rpt_family="A-rich"
repeat_region	17920..17970	/rpt_family="(CAATA)n"
repeat_region	18178..18658	/rpt_family="GA-rich"
repeat_region	18644..18953	/rpt_family="L1"
repeat_region	19026..19155	/rpt_family="L1"
repeat_region	19061..19084	/rpt_family="L2"
repeat_region	19167..19441	/rpt_family="AT-rich"
repeat_region	19638..19749	/rpt_family="L1"
repeat_region	19850..20000	/rpt_family="L1"
repeat_region	20251..20438	/rpt_family="L1"
repeat_region	21123..22815	/rpt_family="MER2_type"
repeat_region	23307..23479	/rpt_family="L1"
repeat_region	24526..24570	/rpt_family="MERL_type"
repeat_region	24570..24849	/rpt_family="(TCTA)n"
repeat_region	25680..25700	/rpt_family="Alu"
repeat_region	26134..26688	/rpt_family="(TA)n"
repeat_region	26944..27375	/rpt_family="ERVL"
repeat_region	27885..28356	/rpt_family="MaLR"
repeat_region	28533..28611	/rpt_family="L2"
repeat_region		/rpt_family="L2"

FEATURES

REFS	source	Location/Qualifiers
		1..159838
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="4"
		/map="4"
		/clone="RP11-512120"
		/clone_lib="RPC1-11"
		2..775
	repeat_region	/rpt_family="MER1_type"
	repeat_region	266..317
	repeat_region	/rpt_family="T-rich"
	repeat_region	770..792
	repeat_region	/rpt_family="AT-rich"
	repeat_region	845..2469
	repeat_region	/rpt_family="L1"
	repeat_region	2470..2766
	repeat_region	/rpt_family="Alu"
	repeat_region	2769..3291
	repeat_region	/rpt_family="MER1_type"
	repeat_region	3321..3441
	repeat_region	/rpt_family="L1"
	repeat_region	3628..3655
	repeat_region	/rpt_family="AT-rich"
	repeat_region	4297..4409
	repeat_region	/rpt_family="MIR"
	repeat_region	4815..5120
	repeat_region	/rpt_family="Alu"
	repeat_region	5094..5114
	repeat_region	/rpt_family="(A)n"
	repeat_region	5203..5546
	repeat_region	/rpt_family="L2"
	repeat_region	8632..8739
	repeat_region	/rpt_family="GA-rich"
	repeat_region	9549..9890
	repeat_region	/rpt_family="ERV1"
	repeat_region	10107..10349
	repeat_region	/rpt_family="MIR"


```

|||||
102220 GGTCTGCTCCAGCACTTGGGAGACCATATATATGAGGAGTATGAGAAAG 102171
17 yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
102170 GAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 102121
34 laAlaGluAsnGluValHisPhePheAspTrpGluHisSerHis 50
102120 CCCCCAGAACAGGACCATCTCTTCGACTCGCAGAGCAATACAGCCAC 102071
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 67
102070 GATTTGGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 102021
67 uThrValGluHisThrProAlaTyrPheThrSerProLysValProGluA 84
102020 CACAGTGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 101971
84 rgValTyrSerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAsp 100
101970 GACTCTACAGCATGAACCGTCCATCCGCTGCTGCTCATCTCGCAGAC 101921
101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
101920 CCGTCGAGAGCGTGATGATGATGATGATGATGATGATGATGATGATGAT 101871
117 tGlnLysHisLysProTyrProSerIleGluGluPheLeuValArgAspG 134
101870 GCAAGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 101821
134 lyArgLeuAsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
101820 GCAGCGTCAATGAGGACTACAGGCGCTCAACCGCAGCCTCTACACCGTG 101771
151 HisMetGluAsnTrpLeuArgPhePheProLeuArgHisIleHisIleVa 167
101770 CATATGAGAACTGGCTGGCTTTTTCGCGCTGCGCCACATCCACATGTT 101721
167 lAspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184
101720 GAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 101671
184 luArgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
101670 AGAGGTTCCTAAAGCTGTCGCCGACATCAATGCTTCCAACTTCTACTTT 101621
201 AsnLysThrLysGlyPheTyrCysLeuAlaAspSerGlyArgAspArgCy 217
101620 AACAAAGAGAGATTTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 101571
217 sleuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeu 234
101570 CTATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 101521
234 euAsnLysLeuHisGluTyrPheHisGluProAsnLysLysPhePheGlu 250
101520 TCAATAAAGTGCAGCAATATTTTCATGAGCCAAATGAAGAAGTCTTCGAG 101471
251 LeuValGlyArgThrPheAspTrpHis 259
101470 CTGTGTCGACAGACATTTGACTGGGCAC 101444

```

seq_name: gb_ro:BC009133

seq_documentation_block:

LOCUS BC009133

1657 bp mRNA linear pop 12-JUL-2001

DEFINITION Mus musculus, heparan sulfate (glucosamine) 3-O-sulfotransferase 1,

clone MGC:11450 IMAGE:3155049, mRNA, complete cds.

ACCESSION BC009133

VERSION BC009133.1

KEYWORDS MGC, GI:14318662

SOURCE house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

Ekaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 1657)

Strausberg, R.
Direct Submission
Submitted (05-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MG), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: mgc-help@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center

Center code: BCM-HGSC
Web site: <http://www.hqsc.bcm.tmc.edu/cdna/>
Contact: villalobosbcm.tmc.edu
Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate 16 Row: 5 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRFA q1 261Rq70.

FEATURES

source

Location/Qualifiers

1..1657

/organism="Mus musculus"

/db_xref="LocusID:15476"

/db_xref="taxon:10090"

/clone="MGC:11450 IMAGE:3155049"

/tissue_type="Mammary tumor, Breast-/fl: MMTV-Cre model. 10

months old, gross tissue."

/clone_lib="NCI_CGAP_Mam3"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

284..1219

/codon_start=-1

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein_id="AAH09133.1"

/db_xref="GI:14318663"

/translation="MTLLLGAVLLVAQPOLVHSHEAAPKGLKQELLKVIILPED

TGESTASNGSTQLPOTIIIGVKGSIPTALLEMLSDHPUVAAAEVHFFUWEIYSQ

GLAWYLCTMPFSSPHGLIVERTPAFTSPRPVPTIISMNPTPLLLILDKPSERVLSD

YTOVLYNHQKHKPYPIEDLMRDGLINIDYKALNRSLYHAHMLNWLRFPLGHIHI

VDSPTLIIPPPPIQKVERFTKISQINASNFYFNKTKGFCYIWNQKIDQCIHESKGR

AHQVDKPLLDLHEYPHEPNKKFKLVGRTFDWH"

BASE COUNT 398 a 439 c 393 g 427 t

ORIGIN

alignment_scores:

Quality: 1318.00 Length: 259

Ratio: 5.169 Gaps: 0

Percent Similarity: 98.456 Percent Identity: 92.278

alignment_block:

US-09-557-262-4_Copy_49_307 x BC009133

Align seg 1/1 to: BC009133 from: 1 to: 1657

1 GlySerAlaGlnLeuProGlnThrIleIleGlyValArgLysG1 17

|||||:|||||

440 GGTTCACACACACAGCGTCCACAGACCATCATCATTTGGCGTGCCGACGGG 489

17 yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34

|||||

```

490 TGGTACCGGAGCGGCTGCTAGACATGAGTGAAGCTTATATGATGATGTTG 530
34  laaIaGluAsnGluValHisPhePheAspTrpGluGluHisTyrSerHis 50
|||||
540 CAGCTCAAAACGAGCTCATTTCTTCTGACTGGGAGGACCAATTACAGCCAA 589
|||||
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||
590 GCGCTGGGCTGCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 639
|||||
67  uhrValGluIysThrProAlaTyrPheThrSerProIysValProGlu 84
|||||
640 CAGCTGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 689
|||||
84  rGValTyrSerMetAsnProSerIleAlaGluLeuLeuIleLeuAsp 100
|||||
690 GAATTCACAGCAATCAAAACCAACCAACCAACCAACCAACCAACCAACCA 739
|||||
101 ProSerGluAraValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
|||||
740 CCATCAGACGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 789
|||||
117 tGlnIysHisIysProTyrProSerIleGluGluPheLeuValArgAspG 134
|||||
790 TCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 839
|||||
134  lyArgLeuAsnValAspTyrIysAlaLeuAsnArgSerLeuTyrHisVal 150
|||||
840 GTCGCTCAACCTGACCTACACAGCTGTCAACCTGACCTGACCTGACCTG 889
|||||
151 HisMetGluAsnTrpLeuArgPhePheProLeuArgHisIleHisIleVa 167
|||||
890 CACATGCTGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 939
|||||
167  IAspGlyAspArgLeuIleArgAspProPhePheGluIleGlnIysValG 184
|||||
940 GATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 989
|||||
184  luArgPheLeuIysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
|||||
990 AAAGATTCCTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1039
|||||
201 AsnLysTrpLysGlyPheTyrCysLeuArgAspSerGlyArgAspArgCy 217
|||||
1040 ACAAACCAACAGCGCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089
|||||
217  sIcuHisGluSerIysGlyArgAlaHisProGlnValAspProIysLeu 234
|||||
1090 CTACACAGTCTCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
|||||
234  cuAsnLysLeuHisGlyTyrPheHisGluTrpAsnLysIysPheGlu 250
|||||
1140 TTGATAAACTCCAGCAATACCTTTCATGCGGCAATAAGCAAAATTTTCA 1189
|||||
251  LeuValGlyArgThrPheAspTrpHis 259
|||||
1190 CTCCTGGGCAACAACTTCGACTGGCAC 1216

seq_name: gb_09-557-262-4_copy_49_307.p2n.rge
seq_documentation_block:
LOCUS      AF019385                1685 bp    mRNA    linear    ROD 14-NOV-1997
DEFINITION Mus musculus heparan sulfate 3-O-sulfotransferase-1
            precursor (3OST1) mRNA, alternatively spliced, complete cds.
ACCESSION  AF019385
VERSION    AF019385.1  GI:2618970
KEYWORDS   3-O-sulfotransferase
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1  (bases 1 to 1685)
AUTHORS   Shworak,N.W., Liu,J., Fritze,L.M.S., Schwartz,J.J., Zhang,L.,
            Loqeart,D. and Rosenberg,R.D.
            Molecular cloning and expression of mouse and human cDNAs encoding
            heparan sulfate D-glucosaminyl 3-O-sulfotransferase
            J. Biol. Chem. 272 (44), 28094-28019 (1997)
            98010647
            2  (bases 1 to 1685)
            Shworak,N.W., Liu,J., Fritze,L.M.S., Schwartz,J.J., Zhang,L.,
            Loqeart,D. and Rosenberg,R.D.
            Direct Submission
            Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
            02139, USA
FEATURES             Location/Qualifiers
     source            1..1685
                        /organism="Mus musculus"
                        /strain="C3H/An"
                        /db_xref="taxon:10090"
                        /cell_line="clonal; 1 cell" line:1TA clone:1A (described in
                        Shworak et. al., 1996, J. Biol. Chem. 271:27071-82)"
                        /note="poly(A)+ RNA isolated from postconfluent cultures"
     5'UTR            1..322
                        /gene="3OST1"
                        /note="alternatively spliced, this version represents a
                        Class 1 mRNA"
     gene              1..1685
     misc_feature      216..217
                        /gene="3OST1"
                        /note="splice variant site; many additional 3-OST-1 cDNAs
                        (Class 2 to Class 9) contain extra 5' UTR exons between
                        these nucleotides (extra 86 to 1629 bp)"
     sig_peptide       323..382
     CDS               323..1258
                        /gene="3OST1"
                        /note="rate limiting enzyme for the synthesis of
                        anticondulant heparan sulfate"
                        /note="heparan sulfate sulfotransferase, 3 OST-1;
                        interluminal Golgi resident protein (retention mechanism
                        unknown)"
                        /codon_start=1
                        /product="heparan sulfate D-glucosaminyl
                        3-O-sulfotransferase-1 precursor"
                        /protein_id="AA084387.1"
                        /db_xref="GI:2618971"
                        /translation="MTLLALGAVLILVAQQLVHSHPAAPGCIKQOFILKKVILILPED
                        TGGTASNGSTQQLPQTIIIGVRKGTGTRALLLEMLSLHDDVAAAEVHFFPWHERVYSO
                        GLGWYLTOMPPSSPHQIVKTIPTVPSKVPKTHSMNPTIRLLLLLRDSRVLSL
                        VTQVLYNHLQKIKPYPIEDLLMPDGRINLDYKALNPSLYIAIMLNWLRPPLPLGHIII
                        VAGDRIRIDPPFEIQRVERFLKLSQINASNYFNKTKGFYLPISGKEPLQLHESKGR
                        AHPQVDPKLLDKLHEYFHEPNKKEFLVGRFTDWH"
                        383..1255
                        /gene="3OST1"
                        /note="confirmed by N-terminal analysis"
                        /evidence="experimental"
                        /product="heparan sulfate D-glucosaminyl
                        3-O-sulfotransferase-1"
                        476..484
                        /gene="3OST1"
                        /note="encodes N-linked glycosylation sites"
                        /evidence="experimental"
                        494..1255
                        /gene="3OST1"
                        /note="encodes presumptive sulfotransferase catalytic
                        domain"
                        908..916
                        /gene="3OST1"
                        /note="encodes N-linked glycosylation sites"
                        /evidence="experimental"
                        1058..1066
                        /gene="3OST1"
                        /note="encodes N-linked glycosylation sites"
                        /evidence="experimental"
                        1079..1087

```

```

/misc_structure /gene="30stl"
/notes="encodes N-linked glycosylation sites"
/evidence=experimental
1100..1129
/gene="30stl"
/notes="encodes cysteine-bridged peptide loop"
/evidence=experimental
1259..1285
/gene="30stl"
1532..1537
/gene="30stl"
1536
/gene="30stl"
1586..1591
/gene="30stl"
1590..1595
/gene="30stl"
1605
/gene="30stl"
1610
/gene="30stl"
1611..1615
/gene="30stl"
1629
/gene="30stl"
1667..1672
/gene="30stl"
1685
/gene="30stl"
BASE COUNT 422 a 426 c 382 g 455 t
ORIGIN

```

```

alignment_scores:
  Quality: 1318.00      Length: 259
  Ratio: 5.169          Gaps: 0
  Percent Similarity: 98.456  Percent Identity: 92.278

alignment_block:
US-09-557-262-4_copy_49_307 x AF019385 ..
Align seq 1/1 to AF019385 from 1 to 1645

1 GlySerAlaGlnGluLeuProGlnThrIleIleGlyValArqLysGI 17
|||||
479 GGTTCACACACACAGCTGCCACACACACACACACACACACACACAGG 528
17 yGlyThrArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||
529 TGGTACCGAGAGCTGTATAGATATATATATATATATATATATATATG 578
34 IaAlaGlnAsnGluValHisPheAspTrpGluGluHisTyrSerHis 50
|||||
579 CAAATGAAAGAGTGTATATATATATATATATATATATATATATATAT 528
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 57
|||||
629 GGGCGGGCTGTACCTACACACACACACACACACACACACACACACAC 678
67 uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
|||||
679 CACCTGTAGAGAGATACAGAGATATATATATATATATATATATATAT 728
84 rGValTyrSerMetAsnProSerIleArgGluLeuLeuIleLeuArqAsp 100
|||||
729 GAATGCAATATGAAAGATATATATATATATATATATATATATATATAT 778
101 ProSerGluArqValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
|||||
779 CATACAGACCGCGCTGTGTGTATATATATATATATATATATATATAT 828
117 tGlnLysHisTyrProTyrProSerIleGluPheLeuValArqAspG 134
|||||

```

```

829 TCAGAAAGAAAGAGCTATCCAAATATAGAAATCTTAATGGGCAAG 878
134 yArqLeuAsnValAspTyrLysAlaIleuAsnArgSerLeuTyrHisVal 150
|||||
879 GTCCGCTCAAGCTGTGACACAGGCTTCAACCCGACCTGTACCAAGCA 928
151 HisMetGlnAsnTrpLeuArqPhePheProLeuArqHisIleHisIleVa 167
|||||
929 CACAAGCTCAACAGCTGCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 978
167 AspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184
|||||
978 GATGTAAGATATATATATATATATATATATATATATATATATATATAT 1028
184 IaArqPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
|||||
1029 AAAGATTCCTCAAGCTTCTCCACAGATCAACGCTCGAAGCTTCTTCT 1078
201 AsnLysThrLysGlyPheTyrCysLeuArqAspSerGlyArqAspArgG 217
|||||
1079 AAAAAAACCAGCGCTTCTACTGCTGTGCGGCACAGTGGCAAGGACCGCT 1128
217 sleuHisGluSerLysGlyArqAlaHisProGlnValAspProLysLeuL 234
|||||
1129 CTATACAGAGATCAAAAGAGCGAGAGAGAGAGAGAGAGAGAGAGAG 1178
234 cuAsnLysHisGlnTyrPheHisGluProAsnLysLysPhePheGlu 250
|||||
1179 TTGATAAACTCCAGCAATACITTCATGAGCCAAATACAAAAATTTTCA 1228
251 LeuValGlyArqThrPheAspTrpHis 259
|||||
1229 CTCGTGGCGAGAACATTCGACTGGCAC 1255
seq_name: gb_htg.AC084416_3
seq_documentation_block:
WPCOMMENT
Sequence split into 6 fragments JACUS AC084416 Accession AC084416
Fragment Name Begin End
AC084416_0 1 110000
AC084416_1 100001 210000
AC084416_2 210001 310000
AC084416_3 310001 410000
AC084416_4 410001 510000
AC084416_5 510001 595289
Continuation (4 of 6) of AC084416 from base 300001 (AC084416 Mus musculus clone rp23)

alignment_scores:
  Quality: 1318.00      Length: 259
  Ratio: 5.169          Gaps: 0
  Percent Similarity: 98.456  Percent Identity: 92.278

alignment_block:
US-09-557-262-4_copy_49_307 x AC084416_3/rev ..
Align seq 1/1 to reverse of: AC084416_3 from 1 to 110000

1 GlySerAlaGlnGluLeuProGlnThrIleIleGlyValArqLysGI 17
|||||
77721 GATTCGAATACAGAGATATATATATATATATATATATATATATAT 77672
17 yGlyThrArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||
77671 TGGTACCGAGAGCTGTATAGATATATATATATATATATATATATAT 77622
34 IaAlaGlnAsnGluValHisPheAspTrpGluGluHisTyrSerHis 50
|||||
77621 CACCTGTAGAGAGATACAGAGATATATATATATATATATATATATAT 77572
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||

```



```

1  GlySerAlaGlnGlnLeuProGlnThrIleIleIleIleIleValArgLysG1 17
|||||...|||||
112077 GCCTCCACACAGCAGCTCCACAGACCATCATTCGGGTGGCCCAAGCG 112076

17  yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||...|||||
112077 TGGTACCCGAGGATATGATAGATATCTTAGCTGATCTCATGTCTGCTG 112126

34  laAlaGluAsnGluValHisPheAspTrpGluGluHisTyrSerHis 50
|||||...|||||
112127 CAGCTGAAAAGAGGATGATCTTCTTCTATCTGAGAGGACATTAACAGCAA 112176

51  GlyLeuGlyTyrTrpLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||...|||||
112177 GRCCTGGGCTGCTACCTCACCACAGATCCCTCTCCCTCCCTCCACAGCT 112226

67  uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
|||||...|||||
112227 CAGCTGGAGAGAGAAAGAGAGGATATTTACTTATGAGAGAGAGAGAG 112276

84  rgValTyrSerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAsp 100
|||||...|||||
112277 GAATCCACAGCATGAACCCAGCATCCAGCTGCTGCTTATCTCAGGAGC 112326

101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
|||||...|||||
112327 CCATCAGAGCGGTGCTCTCCCACTACACCCAGCTGTGTGACACACAGCT 112376

117  tGlnLysHisLysProTyrProSerIleGluGluPheLeuValArgAspG 134
|||||...|||||
112377 TCAGAGAGCAGAGCTTATCAGCATCATCATCATCATCATCATCATCAT 112426

134  lyArgLeuAsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
|||||...|||||
112427 GTGGCTGAACCTGAGATACAGAGCTTTCTAATGAGAGAGAGAGAGAG 112476

151 HisMetGlnAsnTrpLeuArgPhePheProLeuArgHisIleHisIleVa 167
|||||...|||||
112477 CACATGCTGAACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 112526

167  lAspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184
|||||...|||||
112527 GGATGGAGAGCGCTCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112576

184  luArgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
|||||...|||||
112577 AAAGATTCTGAAGCTTTCTCCACAGATCAAGGCTGCTGCTGCTGCTGCT 112626

201 AsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArgAspArgCy 217
|||||...|||||
112627 AACAAACCAAGCGCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 112676

217  sLeuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
|||||...|||||
112677 CTACACAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112726

234  euAsnLysLeuHisGluTyrPheHisGluProAsnLysLysPhePheGlu 250
|||||...|||||
112727 TTGATAAAGCTGACGATATCTTCTATGAGCAGCAGCAGCAGCAGCAGC 112776

251 LeuValGlyArgThrPheAspTrpHis 259
|||||...|||||
112777 CTCGTGGGAGAACATTCGACTGGCAC 112803

seq_name: qb_ro:AF177430
seq_documentation_block:
LOCUS AF177430 936 bp mRNA linear ROD 10-MAY-2001
DEFINITION Rattus norvegicus 3-O-sulfotransferase mRNA, complete cds.
ACCESSION AF177430
VERSION AF177430.1 GI:9957243
KEYWORDS

```

Norway rat.
Rattus norvegicus
Pukaryota; Metazoa; Chordata, Craniata, Vertebrata, Euteleostomi;
Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae;
Rattus.

REFERENCE
1 (bases 1 to 936)
Li, Z. Y., Hirayoshi, K. and Suzuki, Y.
Expression of N-deacetylase/sulfotransferase and
3-O-sulfotransferase in rat alveolar type II cells
Am. J. Physiol. 279 (2), L292-L301 (2000)

JOURNAL
2 (bases 1 to 936)
Li, Z. Y., Hirayoshi, K. and Suzuki, Y.
Direct Submission
Submitted (13-Aug-1999) Department of Ultrastructural Research,
Institute for Frontier Medical Sciences, 53 Shogoin Kawahara-cho,
Sakyo-ku, Kyoto 606 8507, Japan

FEATURES
source
1..936
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/cell_type="type II epithelium"
1..936
/note="3' 5' UTR"
/codon_start=1
/product="3-O-sulfotransferase"
/protein_id="AAC09283.1"
/db_xref="GI:9957244"
/translation="MTLLLLGAVLLVAQLVPSHPAAAGPGLKQGGLLKVVLLPPEP
TREGAATNGSTQQLPOTLIIGVRKGTGALLKMLSDPDAANEVHFFDWEHYSQ
GLGWLTPFSSPQLTVEKTPAYFSPKPEIHSNMPTIRLLILLRDPSEKVLSD
YGVLYNHLLQKHKPYFPIEDILMPDGRNLNVYKALNPSLYAHMLMLWLPFPLGHIH
VDGRPIIDPPPELQKVERFLKSPQINASNPYKTKGFYCLRDGSKDRCLHPSKGR
AHPQVDPKLLDKLHVFREPKNKFKLVGRFTDWH"

BASE COUNT 224 a 292 c 224 g 196 t

ORIGIN

alignment_scores:
Quality: 1304/300 Length: 259
Ratio: 5.133 Gaps: 0
Percent similarity: 98.456 Percent identity: 91.892

alignment_block:
US-09-557-262-4_COPY_49_307 x AF177430 ..
Align seg 1/1 to: AF177430 from: 1 to: 936

1 GlySerAlaGlnGlnLeuProGlnThrIleIleIleIleIleValArgLysG1 17
|||||...|||||
157 GCCTCCACACAGCAGCTCCACAGACCATCATTCGGGTGGCCCAAGCG 206

17 yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||...|||||
207 TGGTACCCGAGGATATGATAGATATCTTAGCTGATCTCATGTCTGCTG 256

34 laAlaGluAsnGluValHisPheAspTrpGluGluHisTyrSerHis 50
|||||...|||||
257 CAGCTGAAAAGAGGATGATCTTCTTCTATCTGAGAGGACATTAACAGCAA 406

51 GlyLeuGlyTyrTrpLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||...|||||
307 GRCCTGGGCTGCTACCTCACCACAGATCCCTCTCCCTCCCTCCACAGCT 356

67 uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
|||||...|||||
357 CAGCTGGAGAGAGAAAGAGAGGATATTTACTTATGAGAGAGAGAGAG 406

84 rgValTyrSerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAsp 100
|||||...|||||
407 GAATCCACAGCATGAACCCAGCATCCAGCTGCTGCTTATCTCAGGAGC 456

101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117

457 CCATCCGAACGGTGGTATCCCATACCTACCTAGCTGTATTATCAATCACTT 506
|||||
117 tGlnLysHisLysProTyrProSerIleGluGluPheLeuValArgAspG 134
|||||
507 TCACAAGCACAAACCTTACCCACCAATTCAGCACTCTCAATGCTGCTGAGG 556
|||||
134 yArgLeuAsnValAspPheTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
|||||
557 GACGGCTCAATGTGATATCAAGGATCTCAACCGCAAGCTGTATATGTA 606
|||||
151 HisMetGluAsnTrpLeuArgPhePheProLeuArgHisIleHisIleVal 167
|||||
607 CACAAGCTCAACTGGCTGGCTTTTCTCCACTGGGCGCACATCCATACCT 656
|||||
167 IaspGlyAspArgLeuLeuArgAspPropheProGluIleGluLysValG 184
|||||
657 GCATGGCCACCGTTTCATCATCAGACAGCTTTTCCCTTCACATCCACAAAGTGG 706
|||||
184 IuArgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPhenylPhe 200
|||||
707 AAGGTGTCTCAAGCTGTCTGCATASATCAACGCTCGAAGCTTCTACTTT 756
|||||
201 AsnLysThrTrpSerGlyPheTrpLysLeuArgAspSerGlyArgAspArgG 217
|||||
757 AACAAACCAAGGCTTCTATACGCTGGGCGCACTGGCGCAAGGACCGCG 806
|||||
217 sLeuIleSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
|||||
807 CTATCATGATCTCAAAAGCGCGCGGCACTCAAGTACATCTCCCAACATC 856
|||||
234 cuAsnLysLeuHisGluTyrPheHisCLeuProAsnLysLysPhePheGlu 250
|||||
857 TTGATAACTGATCAATATTTTGTGTGAGCCCAANTATGAGAAATTTTTCAG 906
|||||
251 LeuValGlyArgThrPheAspTrpHis 259
|||||
907 CTCGTGGGCAGAACTTCGACTGGCAC 933

seq_name: qb_pat:AX317978

seq_documentation_block:	
LOCUS	AX317978
DEFINITION	Sequence 18 from Patent WO0190334.
ACCESSION	AX317978
VERSION	AX317978.1 GI:17900788
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	Yue H., Sanjanwalla M.S., Baughn M.R., Gandhi A.R., Rinq H.Z., Elliot V., Walla N.K., Yang J., Khan F.A., Rangkumar J., Tang Y.T., Hafila A., Lal P., Nguyen D.H., Yao M.G., Lee E.A., Iribouley C.M., Patterson C., Lu Y., Ding L., Bruns C.M. and

alignment_scores:		
Quality:	772.50	Length: 258
Ratio:	3.527	Gaps: 2
Percent Similarity:	84.884	Percent Identity: 53.101

```

alignment_block:
  CS-69-557-262-4_COPY_49_307_x_AX317978  ..
  Align seq 1/1 to: AX317978 from: 1 to: 2115

      4  GtGlnLeuPheGlnThrIleIleIleGlyValArgGlyGlyThrA  20
      |||
      438  CAGACGCTCCCCCAAGGCCATTAICATGGGGTGCAGAAAGGAGCAAG  487
      |||
      20  qAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaGluA  47
      |||
      488  GCGCCCTGCTTTCAATGCTGAACCTTACATCCGGCAGTAGTCAAAAGCCCTC  537
      |||
      37  snGluValHisPhePheAspTrpGluGluHisTyrSerHisGlyLeuGly  53
      |||
      538  AAGAAATCCCACTTTTGTGATAATGATGAGAAATATGTTAAGGGCATTCAG  587
      |||
      54  TrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValGI  70
      |||
      588  TGGTATAGCAAAAGAAGCGCTTTTCTTACCCCTCAGCAAAATCACAAITGA  637
      |||
      70  uLysThrProAlaTyrPheThrSerProLysValProGluArgValTyrS  87
      |||
      638  AAAGAAGTTCAGCAATATATATCAACAGGAGGGTTCACAAAGGCAATACAA  687
      |||
      87  eMetAsnProSerIleArgLeuLeuLeuIleLeuArgAspProSerGlu  104
      |||
      688  AATAGAACTATGCGATCAATCAAAATGTGTAACATGTGCAGGAGCAATTCACA  737
      |||
      104  ArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMetGlnTyrSH  120
      |||
      738  AGAGCTATTATTCATTAATATATATAGTGTCTATACAGTGGCAAGACAGAGA  787
      |||
      120  sLysProTyrProSerIleLeuGluPheLeuValArgAspGly.....A  135
      |||
      788  CAAAACCTTATACAGGTTTCAGAGAGCTGGCCATAGACCCTAATACATCCG  837
      |||
      135  rGluAsnValAspTyrTyrSAlaLeuAsnArgSerLeuTyrHisValHis  151
      |||
      838  AAGTGAAACACAAATACAAAGCAAGTAAAGCAACAGCACTACACCAAACT  887
      |||
      152  MetGlnAsnTrpLeuArgPheProLeuArgHisIleHisIleValAs  168
      |||
      888  CTGCAAGAGGTGGTGTGAATACITTCCAATGAGCAATTTTCATGCTGCGA  947
      |||
      168  polyAspArgLeuIleArgAspPheProGluIleGlnLysValGluA  185
      |||
      938  TCGAGATCGCCCTATACACCAATCTCTGCCACAAACTATGACCTCTGCAGA  987
      |||
      185  rGluLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsn  201
      |||
      988  AGTTCTCAATTTGCTCTCCAAAGGATAAGTCATATACAAATTTATACTTCA  1037
      |||
      202  LysThrLysGlyPheTyrCysLeuArg.....AspSerGlyArgAspArg  217
      |||
      1038  GCTACCAAGAGGGTTTACCTGCTGCCGTAAATATATATCTTAAATAGGTC  1087
      |||
      217  sLeuHisGluSerLysGlyArgAlaHisProGluValAspProLysLeuL  234
      |||
      1088  CCTGGCGGGCAGCAAGGCGGCGATTTCATCCACAGGTGGAGCCCTCTGTCA  1137
      |||
      234  euAsnLysLeuHisGluTyrPheHisGluTrpAsnLysLysPhePheGlu  250
      |||
      1138  TTACTTAAATIGAGCAAAATCTTTCATCTCTTTTAAATCAAAATTTTACAG  1187
      |||
      251  LeuValGlyArgThrPheAspTrp  258
      |||
      1188  ATCACTGGGAGGACATTCGAATCGG  1211

```


DEFINITION Human DNA sequence from clone RP11-112L15 on chromosome 6, complete sequence.

ACCESSION AL355498

VERSION A1355498.10 GI:12141339

KEYWORDS HTG

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Primates, Catarrhini; Hominoidea; Homo.

AUTHORS Hall, R.

TITLE Direct Submission

JOURNAL Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK E-mail enquiries: humquery@sanger.ac.uk

COMMENT requests: clonerequests@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP. Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

RP11-112L15 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/Varpat/homo.htm>

FEATURES

source 1..165549

Location/Qualifiers

1..165549

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP11-112L15"

/clone_lib="RPI1-11.1"

repeat_region 256..573

/note="AluX repeat. matches 1..312 of consensus"

repeat_region 1685..1988

/note="AluJb repeat: matches 1..298 of consensus"

repeat_region 2013..2318

/note="AluSg repeat: matches 1..305 of consensus"

repeat_region 2776..2807

/note="16 copies 2 mer at 84% conserved"

repeat_region 2864..3160

/note="AluSx repeat: matches 1..293 of consensus"

repeat_region 4049..4240

/note="MIR repeat: matches 105..262 of consensus"

repeat_region 4737..4831

/note="L2 repeat: matches 2408..2710 of consensus"

repeat_region 5020..5364

/note="L1pA6 repeat: matches 5798..6143 of consensus"

misc_feature 5187..5197

/note="1372 bases of IS186 transposon (X03123) removed here. This sequence represents the duplicated flanking sequence of the IS186."

repeat_region 5646..6640

/note="L1p4 repeat: matches 4784..5781 of consensus"

repeat_region 7170..7271

/note="34 copies 3 mer tat 61% conserved"

repeat_region 7173..7262

/note="15 copies 6 mer tatata 72% conserved"

repeat_region 7185..7260

/note="38 copies 2 mer ta 77% conserved"

repeat_region 8044..8192

/note="MIR repeat: matches 72..237 of consensus"

repeat_region 8200..8415

/note="MER63B repeat: matches 209..436 of consensus"

repeat_region 8416..8708

/note="AluSg repeat: matches 1..233 of consensus"

repeat_region 8709..8954

/note="MER63A repeat: matches 1..209 of consensus"

repeat_region 8955..9031

/note="MIR repeat: matches 194..241 of consensus"

repeat_region 9130..9173

/note="MIR repeat: matches 76..119 of consensus"

repeat_region 9984..10051

/note="L1MC3 repeat: matches 6740..6807 of consensus"

repeat_region 10112..10472

/note="MIR1A1 repeat: matches 1..365 of consensus"

repeat_region 10538..10734

/note="MER58A repeat: matches 22..224 of consensus"

repeat_region 10747..10856

/note="L1M4 repeat: matches 3667..3737 of consensus"

repeat_region 10985..11176

/note="L2 repeat: matches 2471..2567 of consensus"

repeat_region 11273..11438

/note="AluJb repeat: matches 144..297 of consensus"

repeat_region 11439..11742

/note="AluSx repeat: matches 1..304 of consensus"

repeat_region 11743..11788

/note="AluJb repeat: matches 97..144 of consensus"

repeat_region 11885..11967

/note="L2 repeat: matches 2637..2718 of consensus"

repeat_region 13808..13919

/note="MIR repeat: matches 35..147 of consensus"

repeat_region 14580..14874

/note="AluSx repeat: matches 1..295 of consensus"

repeat_region 15559..15980

/note="match: STS: EM:G60413"

repeat_region 15712..15779

/note="34 copies 2 mer ca 70% conserved"

repeat_region 15716..15775

/note="10 copies 6 mer cacaca 81% conserved"

repeat_region 17139..17258

/note="FLAM_A repeat: matches 1..122 of consensus"

repeat_region 18171..18700

/note="MLT1D repeat: matches 2..505 of consensus"

repeat_region 18863..18939

/note="AluJb/FLAM repeat: matches 43..124 of consensus"

repeat_region 18940..19217

/note="AluJb repeat: matches 1..304 of consensus"

repeat_region 19289..19437

/note="MLT1B repeat: matches 118..264 of consensus"

repeat_region 19558..19694

/note="MLT1B repeat: matches 345..483 of consensus"

repeat_region 20090..21304

/note="Tigger3b repeat: matches 1..1230 of consensus"

repeat_region 22028..22075

/note="L2 repeat: matches 2458..2508 of consensus"

repeat_region 23080..23446

/note="AluSg repeat: matches 41..308 of consensus"

repeat_region 29225..29588

/note="THE1C repeat: matches 1..371 of consensus"

repeat_region 29765..30063

/note="AluSg repeat: matches 1..289 of consensus"

repeat_region 33420..33708

/note="AluSx repeat: matches 1..292 of consensus"

repeat_region 32985..33283

/note="AluSx repeat: matches 1..301 of consensus"

repeat_region 34340..34786

/note="L2 repeat: matches 2348..2703 of consensus"

repeat_region 34858..35006

JOURNAL Anal Biochem 284 (2): 231-239 (2000)
 MEDLINE 20422104
 REFERENCE 2 (bases 1 to 519)
 AUTHORS Christophe D., I P I R H N , n l B., route
 TITLE Direct Submission
 JOURNAL Submitted (23-JUN-1999) Christophe D., I P I R H N , n l B., route
 de Lennik 808, Bldg. C, 1070 Bruxelles, BELGIUM
 COMMENT Related sequence-AF105374
 FEATURES
 Location/Qualifiers
 source
 1..519
 /organism="Canis familiaris"
 /db_xref="taxon:9615"
 /clone="BC1"
 /tissue_type="thyroid"
 1..519
 /gene="3OST1"
 <1..519
 /gene="3OST1"
 /codon_start=3
 /product="heparan sulfate 3-O-sulfotransferase-1
 precursor"
 /protein_id="CA846836.1"
 /db_xref="GI:5441573"
 /db_xref="SPTRMBL:09XST1"
 /translation="AVLLAAQLPVSKVAAPGPAATLQVDESPUGAAPNNSAQQ
 LPQTLIKGPKGTFRALLEMLSHPIVAAAEVHEFFWEHYSGGLGQMPFSS
 PHQTLIKGPKGTFRALLEMLSHPIVAAAEVHEFFWEHYSGGLGQMPFSS
 PYPSEHFLVRD"
 BASE COUNT 81 a 199 c 169 g 70 t
 ORIGIN

alignment_scores:
 Quality: 667.00 Length: 133
 Ratio: 5.053 Gaps: 0
 Percent Similarity: 99.248 Percent Identity: 94.747

alignment_block:
 US-09-557-262-4_COPY_49_307 x CFA388534 ..

Align seq 1/1 to: CFA388534 from 1 to: 519

1 GlySerAlaGlnGlnProGlnThrIleTleGlyValAlaGlyLysL 17
 |||||
 120 GGTCTGGCCGACGAGCTGCGGACAGCAACATCGGGGTGGCAAGG 169
 17 YGlyThrArgAlaLeuLeuGluMetIleuSerIleuHisProAspValAla 34
 |||||
 170 CGGCAAGCGGCGTGTGTGATGTCAGCTGACGCGACGTGGG 214
 34 IaAlaGluAsnGluValHisPheProAspTTPGluGluHisTyrSerHis 50
 |||||
 220 CCGCCGAGAACGAGTTCCTCTTCGAGTGGGAGGACACTACGACG 269
 51 GlyLeuGlyTrpTyrIleuSerGlnMetProPheSerTrpProHisGlnLe 67
 |||||
 270 GGCTGGGTGGTCTACCTGGCCAGATGCCCTTCCTCCCGCCACGCT 319
 67 uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
 |||||
 320 CAGGTGGAATAGAGAGAGAGAGATTCAGTGGGAAAATGAGTGG 369
 84 rgValTyrSerMetAsnProSerIleAlaPheLeuLeuLeuAlaAsp 100
 |||||
 370 GGTGTCACAGATAGAACCGGAGATCTGGTGTGATGATGAGTGG 419
 101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
 |||||
 420 CCGTCGAGCGGCTGCTGCGACATACACCAAGGCTCTACACACGCT 469
 117 tGlnIlyHisLysProTyrTrpSerIleGluPheLeuValArgAsp 133
 |||||
 470 GCAAGAGCGCAAGAGATAGAGTGTATGAGAGATTTGATGAGAG 518

seq_name: qb_pr:AF105374
 seq_documentation_block:
 LOCUS AF105374 1968 bp mRNA linear PRI 11-JUN-2001
 DEFINITION Homo sapiens heparan sulfate D-glucosaminyl 3-O-sulfotransferase 2
 (3OST2) mRNA, 3OST2-1 allele, complete cds.
 ACCESSION AF105374
 VERSION AF105374.1 GI:4835718
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1968)
 AUTHORS Shworak N.W., Liu J., Petros I.M., Zhang L., Kobayashi, M.,
 Cupciand N.G., Jenkins, N.A. and Rosenberq, R.D.
 TITLE Multiple isoforms of heparan sulfate D-glucosaminyl
 3-O-sulfotransferase. Isolation, characterization, and expression
 of human cDNAs and identification of distinct genomic loci
 J. Biol. Chem. 274 (8): 5170-5184 (1999)
 JOURNAL 99143187
 MEDLINE 2 (bases 1 to 1968)
 REFERENCE Shworak N.W., Liu, J. and Rosenberq, R.D.
 AUTHORS Direct Submission
 TITLE Submitted (12-NOV-1998) Biology, Massachusetts Institute of
 Technology, 31 Ames St, Cambridge, MA 02139, USA
 JOURNAL location/Qualifiers
 FEATURES
 source
 1..1968
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p12"
 /tissue_type="brain"
 /note="isolated by hybridization screening using IMAGE
 Consortium clone c 20d10 found in GenBank Accession Number
 F07258"
 1..1968
 /gene="3OST2"
 /allele="3OST2-1"
 <1..172
 /gene="3OST2"
 73..1176
 /gene="3OST2"
 /EC_number="2.8.2.23"
 /note="3-OST-2"
 /codon_start=1
 /product="heparan sulfate D-glucosaminyl
 3-O-sulfotransferase 2"
 /protein_id="AA830206.1"
 /db_xref="GI:4835719"
 /translation="MAYRVILGRAGPPQPRRRRLIPFTLSISCTVLCYSFICCCDDL
 GPSNIIIGAPPCILGPGAGGOKLQKSRGCDPSGPTTSPSPASATAAVAAPRIKSGN
 HSGSPKLTGKRLPQALTIVGKGGTAVLEFIVPHVPALATEPHFDNRYGGLDW
 YRSLMPRTLESQITLKTSTPSYFTQEPARIFNMGRDTKLIIVVKNPVTIAISDYTOT
 LSKKPDPTFEGISFRNRTGLGVDSNNAIRIGVMVHLESLMLOYEPLAQIHFVSSER
 LITDPAGMCRVQDFICIKRFTIDKHFFPNKIKGFCCLKAKTESLILPCLCKSKGRTH
 VQIDPRVIDLREFYRPNYKIPYETVCGDFRWF"
 73..129
 /gene="3OST2"
 /note="cytoplasmic domain; unclassified site"
 130..195
 /gene="3OST2"
 /note="transmembrane-region site"
 196..199
 /gene="3OST2"
 /note="SPLAG domain; Ser, Pro, Leu, Ala, Gly enriched;
 unclassified site"
 235..291
 /gene="3OST2"
 /note="putative cystine bridged peptide loop, unclassified
 site"
 307..309

```

/misc_feature /gene-"3OST2"
313..315 /note-"putative O-glycosylation site"
/gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
322..324 /gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
331..333 /gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
367..375 /gene-"3OST2"
/misc_feature /note-"putative N-glycosylation site"
400..417 /gene-"3OST2"
424..462 /note-"sulfolransferase domain, unclassified site"
/gene-"3OST2"
467..714 /note-"putative N-glycosylation site"
667..714 /note-"putative sheet-loop-helix structure for binding the
3'-phosphate of PAPS; binding site"
730..732 /gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
775..783 /gene-"3OST2"
804 /note-"putative N-glycosylation site"
/misc_feature /gene-"3OST2"
823..852 /note-"3OST2-2 allele, silent wobble position mutation"
/replace-"a"
/gene-"3OST2"
852..896 /note-"putative structure for binding 5'-sulfate of PAPS;
binding site"
988..996 /gene-"3OST2"
/misc_feature /note-"putative N-glycosylation site"
1009..1047 /gene-"3OST2"
/misc_feature /note-"putative cystine bridged peptide loop; unclassified
site"
1102..1110 /gene-"3OST2"
/misc_feature /note-"putative N-glycosylation site"
1177..11950 /gene-"3OST2"
1249 /gene-"3OST2"
/misc_feature /note-"3OST2-2 allele"
/replace-"q"
1350 /gene-"3OST2"
1350 /note-"3OST2-2 allele"
/replace-"c"
1507 /gene-"3OST2"
1743..1748 /note-"3OST2-2 allele"
/replace-"t"
/misc_feature /gene-"3OST2"
1925..1930 /note-"putative"
/gene-"3OST2"
1950

```

```

BASE COUNT 441 a 603 c 490 g 434 t
ORIGIN
alignment_scores:
  Quality: 598.00 Length: 261
  Ratio: 2.990 Gaps: 4
  Percent Similarity: 76.628 Percent Identity: 44.295
alignment_block:
  US-09-557-262-4_COPY_49_307 x AF105374
  Allan seq 1/1 to: AF105374 from: 1 to: 1968
4 GlnGlnLeuProGlnThrIleIleIleGlyValAlaGlySerGlyGlyThr 20
406 AAGGCGTTGGCGCAAGAGGCTGATTGTGGGGTGAAGAGGGGGGCAACGG 455
20 qAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaAlaAla 37
456 GATCTGTGTGGAGTTATCCACACACCGGACGTCGGCCCTTGGGCA 505
37 snGluValHisPhePheAspTrpGluGluHisTyrSerHisGlyLeuGly 53
506 CGGAACCTCACTTCTTGAC AGCAACATACGGCGGGGCTGAT 549
54 TrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValGI 70
550 TGGTATACAGAGCTGATGCGAGAGAGGCTGAGAGAGAAAGATACAGTGG 599
70 TrpThrProAlaTyrPheThrSerProLysValProGluArqValTyrS 87
600 CAACACCGCCAGCTACTTGGACACACAGAGAGGCTGCGACGACATGCA 649
87 erMetAsnProSerIleArqLeuLeuLeuLeuArqAspProSerGlu 103
655 ACATCTTCCTGAGACACAGAGATGATCTGGTGTGGAGAGAACTGTGAG 659
104 ArgValLeuSerAspTyrThrGluValPheTyrAsnHisMetGlnTyrSH 120
700 CGAGCAATCTGTATTACAGACAGACA.....CTGCGCAAGAA 737
120 sLysProTyrProSerIleGluGluPheLeuValArqAsp.....G 134
738 GCGGACATATCCGACATCTTGAAGGCTCTCTCTCCGCAAGCGGACCG 787
134 LyArqLeuAsnValAspTyrTyrLysAlaLeuAsnArqSerLeuTyrHisVal 150
786 GATGTGAGAGAGTGTGTGAGAGAGATGGATATGGGATGATAGTGGCTG 847
151 HisMetGlnAsnTrpGluArqPhePheProLeuArgHisIleHisIleVal 167
838 CAGCTGGAGAGATGGATGTGAGTATCTGAGCTTAACTGATGATCTGATCT 887
167 lAspGlyAspArqLeuLeuArqAspProPheProGluIleGlnLysVal 184
888 CAGGATACAGAGATCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 937
184 LuArqPheLeuTyrLeuSerProGlnIleAsnAlaSerAspPheTyrPhe 200
938 AGGATTTCTGGGATTAAAGATTCATGAGAGAGAGAGAGAGAGAGAGAG 987
201 AsnLysThrLysGlyPheTyrCysteuaArg.....AspSerGlyAr 214
988 AACAGAGACAAAGGATTCCTGTTGAAAAAAGAGAGAGAGAGAGAGAGAG 1037
214 qAspArqCysLeuHisGluSerLysGlyArqAlaHisProGlnValAsp 231
1038 GCGTGGATCTTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1087
231 rOLysLeuLeuAsnLysLeuHisGluTyrPheHisGluProAsnLysLys 247

```

```

1088 CTTAAGTGTATACACAGTTCGTCATTAATTTATAGACGCTATATATCAAA 1117
248 PhePheGluLeuValGluArgThrPheAspTrp 258
111:::111 1111111:::111 111
1138 TTTTATGAACCGTTGGCAGGATTTTCAGTGG 1170

seq_name: db_pt.AF105375
seq_documentation_block:
LOCUS AF105375 1968 bp mRNA linear PFI 15-OCT-2001
DEFINITION Homo sapiens heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2
(3OST2) mRNA, 3OST2 2 allele, complete cds.
ACCESSION AF105375
VERSION AF105375.1 GI:4835720
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Claniata; Vertebrata; Euteleostomi,
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1968)
Shworak,N.W., Liu,J., Petros,L.M., Zhang,L., Kobayashi,M.,
Copeland,N.G., Jenkins,N.A. and Rosenberg,P.D.
Multiple isoforms of heparan sulfate D-glucosaminyl
3-O-sulfotransferase. Isolation, characterization, and expression
of human cDNAs and identification of distinct genomic loci
J. Biol. Chem. 274 (8), 5170-5184 (1999)
JOURNAL 99143187
MEDLINE 2 (bases 1 to 1968)
Shworak,N.W., Liu,J. and Rosenberg,P.D.
AUTHORS Direct Submission
TITLE Submitted (12-NOV-1998) Biology, Massachusetts Institute of
JOURNAL Technology, 31 Ames St, Cambridge, MA 02139, USA
FEATURES
location/qualifiers
1..1968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p12"
/issue_type="brain"
/notes="Isolated by hybridization screening using IMAGE
Consortium clone c-20d10 found in GenBank Accession Number
F07258"
gene 1..1968
/gene="3OST2"
/allele="3OST2-2"
1..72
/gene="3OST2"
73..1176
/gene="3OST2"
/PC_number="2.8.2.23"
/notes="3-OST-2"
/codon_start=1
/product="heparan sulfate D-glucosaminyl
3-O-sulfotransferase-2"
/protein_id="AAB30207.1"
/db_xref="GI:4835721"
/translation="MAYRVLGKAGPPEPPAPRLIPAPTLISLSTYLYNSELVUTDL
GRSPLLAGPRLRPAGGKGLKSPPTSPSPSEPSPAPSAFAAAPPPLSGSN
HSSGPKLGLTQALIVGKGGTRAVLEFIVHIDVPALATEPIFFDNNYGPGLDW
YPSLMPPTLRSQITLTKTPSYFVTOQAPRPFNMSPTKLIIVVKNPVTALISDYTOT
LSKKPDIPTFPGISFNRNLTGLVDVSNNAIRICMYVLHLESMILOYPPLAOHPFVSCRR
LITDPAGCMGRVODFLGKIFETIDKHFFYKPKGPKLCKKTKSSLLPRCLGSKGRTH
VQIDPEVIDQLRFYPPYNTKFTYVQDPPWE"
73..129
/misc_feature
/gene="3OST2"
/notes="cytoplasmic domain; unclassified site"
130..195
/misc_feature
/gene="3OST2"
/notes="transmembrane-region site"
196..399
/misc_feature
/gene="3OST2"
/notes="SPIAG domain; Ser, Pro, Leu, Ala, Gly enriched;
unclassified site"
235..291
/misc_feature
/gene="3OST2"
/notes="putative cystine bridged peptide loop; unclassified
site"
298..300
/misc_feature
/gene="3OST2"
/notes="putative O-glycosylation site"
307..309
/misc_feature
/gene="3OST2"
/notes="putative O-glycosylation site"
313..315
/misc_feature
/gene="3OST2"
/notes="putative O-glycosylation site"
322..324
/misc_feature
/gene="3OST2"
/notes="putative O-glycosylation site"
331..333
/misc_feature
/gene="3OST2"
/notes="putative O-glycosylation site"
367..375
/misc_feature
/gene="3OST2"
/notes="putative N-glycosylation site"
400..1173
/misc_feature
/gene="3OST2"
/notes="sulfotransferase domain; unclassified site"
421..462
/misc_feature
/gene="3OST2"
/notes="putative beta-strand/alpha-helix motif;
binds 5'-phosphate of 3' phosphoadenosine 5'-phosphate,
PAPS; binding site"
442..444
/misc_feature
/gene="3OST2"
/notes="catalytic lysine residue; active site"
454..483
/misc_feature
/gene="3OST2"
/notes="putative structure for binding 5'-sulfate of PAPS;
binding site"
640..678
/misc_feature
/gene="3OST2"
/notes="putative cystine bridged peptide loop; unclassified
site"
649..657
/misc_feature
/gene="3OST2"
/notes="putative N-glycosylation site"
667..714
/misc_feature
/gene="3OST2"
/notes="putative sheet-loop-helix structure for binding
3'-phosphate of PAPS; binding site"
730..732
/misc_feature
/gene="3OST2"
/notes="putative O-glycosylation site"
775..783
/misc_feature
/gene="3OST2"
/notes="putative N-glycosylation site"
804
/variation
/gene="3OST2"
/notes="3OST2-1 allele; silent wobble position mutation"
823..852
/misc_feature
/gene="3OST2"
/notes="putative structure for binding 5'-sulfate of PAPS;
binding site"
892..924
/misc_feature
/gene="3OST2"
/notes="putative structure for binding 5'-sulfate of PAPS;
binding site"
937..969
/misc_feature
/gene="3OST2"
/notes="putative structure for binding 5'-sulfate of PAPS;
binding site"
988..996
/misc_feature
/gene="3OST2"
/notes="putative N-glycosylation site"

```


100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_246_272 x ABA50047/rev

Align seq 1/1 to reverse of: ABA50047 from: 1 to: 924

```

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArqAspSerGlyAr 17
|||||
170 TTTCTACTTTTACAAACCAAGGCTTTTACGCTCGGACACAGCGCGG 121
|||||
17 qAspArqCysLeuHisGlySerLysGlyArq 27
|||||
120 GGAAGAGTGTATTATATGATGTTAAAGAGGAG 90

```

seq_name /SINSS/gcgdata/gcg-seq/gcg-seq-emb1/NA2001A.DAT.ABA67978

seq_documentation_block:

ID ABA67978 standard; DNA; 924 BP.

```

XX ABA67978;
XX
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #16283
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX W0200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-0500666.
XX
XX 04-FEB-2000; 2000RS 0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234487.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB 0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI, 2001-483447/53.
XX
XX Human genome derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX
XX Claim 4, SEQ ID NO 16283, 637pp, sequence listing, English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other.

```

alignment_scores:

Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_246_272 x ABA67978/rev

Align seq 1/1 to reverse of: ABA67978 from: 1 to: 924

```

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArqAspSerGlyAr 17
|||||
170 TTTCTACTTTTACAAACCAAGGCTTTTACGCTCGGACACAGCGCGG 121
|||||
17 qAspArqCysLeuHisGlySerLysGlyArq 27
|||||
120 GGAAGAGTGTATTATATGATGTTAAAGAGGAG 90

```

seq_name /SINSS/gcgdata/gcg-seq/gcg-seq-emb1/NA2001A.DAT.ABA35017

seq_documentation_block:

ID ABA35017 standard; DNA; 924 BP.

```

XX ABA35017;
XX
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #13483 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
XX Homo sapiens.
XX
XX W0200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-0500666.
XX
XX 04-FEB-2000; 2000RS 0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234487.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB 0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI, 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 4, SEQ ID NO 13483, 536pp, English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other.

```

alignment_scores:

Quality: 155.00 Length: 27

Align seq 1/1 to reverse of: AAK42122 from: 1 to: 924

```

1 PheYrPheAsnLysThrLysGlyPheCysLeuArgAspSerGlyAr 17
|||||
170 TTCTACTTTTAAACAAACCAAGGGCTTTTACTGCTCGGACAGCGCGG 121

```

```

17 qAspArgCysLeuIleGlySerLysGlyArq 27
|||||
120 GGACCGTGGCTTACATACAGTCCAAAGGCGG 90

```

seq_name: /SDS5/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA148186

seq_documentation_block:

ID_AA148186 standard; DNA: 924 BP.

XX

AC AA148186;

XX 17-OCT-2001 (first entry)

XX DE Probe #16872 used to measure gene expression in human placenta sample

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder; ss.

XX OS Homo sapiens.

XX PN W0200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001W0-NS00663

XX PR 04-PEP-2000; 2000NS-0180332.

XX PR 24-MAY-2000; 2000NS-0207456.

XX PR 30-JUN-2000; 2000NS-0608408.

XX PR 03-AUG-2000; 2000NS-0632356.

XX PR 21-SEP-2000; 2000NS-0234687.

XX PR 27-SEP-2000; 2000NS-0236359.

XX PR 04-OCT-2000; 2000NS-0024263

XX PA (MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PS analyzing gene expression in human placenta -

XX claim 25; SEQ ID No 16872; 654pp; English.

XX CC the present invention relates to single exon nucleic acid probes (SEQ).

XX CC the present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 924 BP; 169 A, 252 C, 310 G, 193 T, 0 other;

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 27

Gaps: 0

US-09-557-262-4_copy_246_272 x AA148186/rev

Align seq 1/1 to reverse of: AA148186 from: 1 to: 924

```

1 PheYrPheAsnLysThrLysGlyPheCysLeuArgAspSerGlyAr 17
|||||
170 TTCTACTTTTAAACAAACCAAGGGCTTTTACTGCTCGGACAGCGCGG 121

```

```

17 qAspArgCysLeuIleGlySerLysGlyArq 27
|||||
120 GGACCGTGGCTTACATACAGTCCAAAGGCGG 90

```

seq_name: /SDS5/gcdata/geneseq/geneseqn-emb1/NA1999.DAT:AAX17246

seq_documentation_block:

ID_AA37246 standard; DNA: 1305 BP.

XX

AC AA37246;

XX 20-JUL-1999 (first entry)

XX DE Human 3-OST-1 encoding DNA.

XX

XX KW Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;

XX KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;

XX KW heparan sulfate; thrombotic disorder; deep vein thrombosis;

XX KW pulmonary embolism; coagulation enzyme inactivation; ss.

XX OS Homo sapiens.

XX PN W09922005-A2.

XX PD 06-MAY-1999.

XX PF 23-OCT-1998; 98WO-US22597.

XX PR 31-OCT-1997; 97US-0065437.

XX PR 24-OCT-1997; 97US-0062762.

XX PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX PI Fritze LMS, Liu J, Rosenberg MB, Schwartz JJ, Shworak NW;

XX PI Zhang L;

XX PR WPI; 1999 312968/26.

XX PR PFSDB, AAV17063.

XX PT Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related

XX PT polynucleotides

XX PS Claim 5; Page 74-75; 95pp; English.

XX CC the invention relates to nucleic acid molecules (AA37245-X17250)

XX CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferase (3-OSTs).

XX CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues

XX CC within a preparation of glycosaminoglycan or proteoglycan

XX CC polysaccharides. 3-OST-1 can be used for enriching the

XX CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).

XX CC 3-OST 1 can also convert HS proteoglycan anticoagulant precursor to the

XX CC HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic

XX CC function) can be used to determine partial sequence information for

XX CC complex polysaccharides. The 3-OST proteins, genes and antibodies are

XX CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs

XX CC are especially used to generate anticoagulant pentasaccharides, which may

XX CC be used to treat thrombotic disorders such as deep vein thrombosis and

XX CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is

XX CC enhanced by complexing of antithrombin with endothelial cell surface HS

XX CC proteoglycans. This is responsible for the non-thrombotic properties

XX CC of blood vessels. The present sequence represents a human 3-OST-1

XX CC encoding DNA.

XX SQ Sequence 1305 BP; 304 A; 380 C; 129 G; 292 T; 0 other;

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 27

Gaps: 0

alignment_block:

US-09-557-262-4_COPY_246_272 x AAX37246

Align seg 1/1 to: AAX37246 from: 1 to: 1305

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyIar 17
 |||||
 854 TTCTATTTTAA'AAACAAACAAAGGCTTTTACAGCTCTGGGACACAGGCGG 903

17 gAspArgCysLeuHisGluSerLysGlyArg 27

|||||

904 GCACCGCTGCTTACATGAGTCCAAAGGCGG 934

seq_name: /SIBS5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAZ37246

seq_documentation_block:

ID AAZ36687 standard; DNA; 1305 BP.

XX AAZ36687;

XX 13-MAR-2000 (first entry)

XX Nucleotide sequence for Genbank accession number AF019386.

XX Stimulus-regulated nucleic acid; sequence profile; nucleic acid level;
 KW differentially expressed nucleic acid; disease state; cancer;
 KW autoimmune disease; infectious disease; aging; developmental disorder;
 KW proliferative disorder; neurological disorder; toxicity;
 KW treatment resistance; differential expression; drug discovery;
 KW growth factor; epidermal growth factor; radiation; stress; pathogen; ss.

OS Homo sapiens.

XX W09955913-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99W0-0509119

XX 27-APR-1998; 990S-0084441.

XX 27-AUG-1998; 980S-0098070.

XX 04-FEB-1999; 990S-0118624

XX (KIMM-) KIMMEL CANCER CENT STINEY

XX McClelland M, Welsh J, Trenkle T;

XX WPI; 2000-086388/07.

XX Measuring expression of low abundance reduced complexity target nucleic

XX acid molecules

XX Disclosure, Fig 14; 187pp, English.

XX AAZ36681-236725 represent nucleic acid sequences from stimulus-regulated

XX nucleic acid molecules. The sequences represent a profile of sequences

XX which can function as targets in the method of the invention. The

XX specification describes a method for measuring the level of two or more

XX nucleic acid molecules in a target. The method comprises contacting a

XX probe with an arbitrarily or statistically sampled target and detecting the

XX amount of specific binding of the target to the probe. The methods

XX can be used to identify differentially expressed nucleic acid molecules

XX associated with disease states, such as cancer, autoimmune disease,

XX infectious disease, aging, developmental disorder, proliferative

XX disorder or neurological disorder. Alternatively, the methods can be

XX used to assess the efficacy or toxicity of or a resistance to a

XX treatment. Also the methods can be used to determine differential

XX expression of nucleic acid molecules in response to a stimulus, e.g. a

XX chemical, drug or growth factor (especially epidermal growth factor),

XX radiation, stress or a pathogen. The methods can also be used to

XX determine co regulated genes that can be potential targets for drug

XX discovery.

XX sequence 1305 BP; 304 A; 686 C; 329 G; 292 T; 9 other;

alignment_scores:

Quality: 155.00 length: 27
 Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_246_272 x AAZ36687

Align seg 1/1 to: AAZ36687 from: 1 to: 1305

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyIar 17
 |||||
 854 TTCTATTTTAA'AAACAAACAAAGGCTTTTACAGCTCTGGGACACAGGCGG 903

17 gAspArgCysLeuHisGluSerLysGlyArg 27

|||||

904 GCACCGCTGCTTACATGAGTCCAAAGGCGG 934

seq_name: /SIBS5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ37245

seq_documentation_block:

ID AAX37245 standard; DNA; 1685 BP.

XX AAX37245;

XX 20-JUL-1999 (first entry)

XX Mouse 3-OST-1 encoding DNA.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;

XX saccharide, glycosaminoglycan, proteoglycan; antithrombin; 3-OST;

XX heparan sulfate; thrombotic disorder; deep vein thrombosis;

XX pulmonary embolism; coagulation enzyme inactivation; ss.

XX Mus musculus.

XX W09955905 A2.

XX 06-MAY-1999.

XX 23-OCT-1998; 98W0-US22597.

XX 31-OCT-1997; 970S-0065437.

XX 24-OCT-1997; 970S-0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Fritz LMS, Liu J, Rosenberg BD, Schwartz JJ, Shworak NW;

XX Zhang L;

XX WPI; 1999-312968/26.

XX P-PSDB; AAY17062.

XX Heparan sulfate D glucosaminyl 3 O sulfotransferase and related

XX polynucleotides

XX Claim 5; Page 71-72; 95pp; English.

XX The invention relates to nucleic acid molecules (AAZ37245, X47250)

XX encoding heparan sulfate D-glucosaminyl 3-O-sulfating saccharide residues

XX within a preparation of glycosaminoglycan or proteoglycan

XX polysaccharides. 3-OST-1 can be used for enriching the

XX antithrombin binding fraction in a preparation of heparan sulfates (HS).

XX 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the

XX HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic

XX function) can be used to determine partial sequence information for

XX complex polysaccharides. The 3 OST proteins, genes and antibodies are

XX also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs

XX are especially used to generate anticoagulant pentasaccharides, which may

XX be used to treat thrombotic disorders such as deep vein thrombosis and

Sequence 3484 BP, 959 A, 844 C, 725 G, 956 I, 0 other,

alignment_scores.

Quality: 114.00 Length: 27
Ratio: 4.560 Caps: 0
Percent Similarity: 92.593 Percent Identity: 74.074

alignment_block.

US-09-557-262-4_copy_246_272 x ABL26622/rev

Align seg 1/1 to reverse of: ABL26622 from: 1 to: 3484

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
1195 TTTTACTTCAATGACCAAGGCTTATATCTGTGTGTATATCAACAG 1146

17 gAspArgCysLeuHisGluSerLysClyArg 27

|||||
1145 GATCGCTGCTCCGCGGACAGCAAGCGCAGG 1115

seq_name: /SIDS53/cgdata/geneseq/geneseqn emb1/NA2001A.DAT.AAS4207

seq_documentation_block.

ID AAS34207 standard; cDNA; 504 BP

XX AAS34207;

XX DT 17-DEC-2001 (first entry)

XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 731.

XX KW Immunomodulator; cardiovascular; ss; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytostatic; nephroprotective;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection.
KW wound healing; epithelial cell proliferation; food additive.

XX OS Homo sapiens.

XX PN WO200155312-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01321.

XX PF 31-JAN-2000; 2000US-0175065.

XX PF 04-FEB-2000; 2000US-0180638.

XX PF 24-FEB-2000; 2000US-0184664.

XX PF 02-MAR-2000; 2000US-0186350.

XX PF 16-MAR-2000; 2000US-0189874.

XX PF 17-MAR-2000; 2000US-0190076.

XX PF 18-APR-2000; 2000US-0198123.

XX PF 19-MAY-2000; 2000US-0205515.

XX PF 07-JUN-2000; 2000US-0209467.

XX PF 28-JUN-2000; 2000US-0214886.

XX PF 30-JUN-2000; 2000US-0215135.

XX PF 07-JUL-2000; 2000US-0216647.

XX PF 07-JUL-2000; 2000US-0216880.

XX PF 11-JUL-2000; 2000US-0217487.

XX PF 11-JUL-2000; 2000US-0217496.

XX PF 14-JUL-2000; 2000US-0218290.

XX PF 26-JUL-2000; 2000US-0220963.
XX PF 26-JUL-2000; 2000US-0220964.
XX PF 14-AUG-2000; 2000US-0224518.
XX PF 14-AUG-2000; 2000US-0224519.
XX PF 14-AUG-2000; 2000US-0225213.
XX PF 14-AUG-2000; 2000US-0225214.
XX PF 14-AUG-2000; 2000US-0225266.
XX PF 14-AUG-2000; 2000US-0225267.
XX PF 14-AUG-2000; 2000US-0225268.
XX PF 14-AUG-2000; 2000US-0225270.
XX PF 14-AUG-2000; 2000US-0225447.
XX PF 14-AUG-2000; 2000US-0225757.
XX PF 14-AUG-2000; 2000US-0225758.
XX PF 14-AUG-2000; 2000US-0225759.
XX PF 14-AUG-2000; 2000US-0226279.
XX PF 24-AUG-2000; 2000US-0226681.
XX PF 24-AUG-2000; 2000US-0226868.
XX PF 22-AUG-2000; 2000US-0227182.
XX PF 23-AUG-2000; 2000US-0227009.
XX PF 30-AUG-2000; 2000US-0228924.
XX PF 01-SEP-2000; 2000US-0229287.
XX PF 01-SEP-2000; 2000US-0229343.
XX PF 01-SEP-2000; 2000US-0229344.
XX PF 01-SEP-2000; 2000US-0229345.
XX PF 05-SEP-2000; 2000US-0229509.
XX PF 05-SEP-2000; 2000US-0229513.
XX PF 06-SEP-2000; 2000US-0230437.
XX PF 06-SEP-2000; 2000US-0230438.
XX PF 08-SEP-2000; 2000US-0231242.
XX PF 08-SEP-2000; 2000US-0231243.
XX PF 08-SEP-2000; 2000US-0231244.
XX PF 08-SEP-2000; 2000US-0231414.
XX PF 08-SEP-2000; 2000US-0231414.
XX PF 08-SEP-2000; 2000US-0232080.
XX PF 08-SEP-2000; 2000US-0232081.
XX PF 12-SEP-2000; 2000US-0231968.
XX PF 14-SEP-2000; 2000US-0232397.
XX PF 14-SEP-2000; 2000US-0232398.
XX PF 14-SEP-2000; 2000US-0232399.
XX PF 14-SEP-2000; 2000US-0232400.
XX PF 14-SEP-2000; 2000US-0232401.
XX PF 14-SEP-2000; 2000US-0233063.
XX PF 14-SEP-2000; 2000US-0233064.
XX PF 14-SEP-2000; 2000US-0233065.
XX PF 21-SEP-2000; 2000US-0234223.
XX PF 21-SEP-2000; 2000US-0234274.
XX PF 25-SEP-2000; 2000US-0234497.
XX PF 25-SEP-2000; 2000US-0234498.
XX PF 26-SEP-2000; 2000US-0235484.
XX PF 27-SEP-2000; 2000US-0235834.
XX PF 27-SEP-2000; 2000US-0235835.
XX PF 29-SEP-2000; 2000US-0236327.
XX PF 29-SEP-2000; 2000US-0236367.
XX PF 29-SEP-2000; 2000US-0236368.
XX PF 29-SEP-2000; 2000US-0236369.
XX PF 29-SEP-2000; 2000US-0236370.
XX PF 02-OCT-2000; 2000US-0236802.
XX PF 02-OCT-2000; 2000US-0237037.
XX PF 02-OCT-2000; 2000US-0237038.
XX PF 02-OCT-2000; 2000US-0237039.
XX PF 02-OCT-2000; 2000US-0237040.
XX PF 13-OCT-2000; 2000US-0239935.
XX PF 13-OCT-2000; 2000US-0239937.
XX PF 20-OCT-2000; 2000US-0240960.
XX PF 20-OCT-2000; 2000US-0241221.
XX PF 20-OCT-2000; 2000US-0241785.
XX PF 20-OCT-2000; 2000US-0241786.
XX PF 20-OCT-2000; 2000US-0241787.
XX PF 20-OCT-2000; 2000US-0241808.
XX PF 20-OCT-2000; 2000US-0241809.
XX PF 20-OCT-2000; 2000US-0241856.
XX PF 01-NOV-2000; 2000US-0244617.
XX PF 08-NOV-2000; 2000US-0246474.
XX PF 08-NOV-2000; 2000US-0246475.
XX PF 08-NOV-2000; 2000US-0246476.
XX PF 08-NOV-2000; 2000US-0246477.
XX PF 08-NOV-2000; 2000US-0246478.
XX PF 08-NOV-2000; 2000US-0246523.
XX PF 08-NOV-2000; 2000US-0246524.
XX PF 08-NOV-2000; 2000US-0246525.
XX PF 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000: 2000US-0244527
 PR 08-NOV-2000: 2000US-0244528
 PR 08-NOV-2000: 2000US-0244532
 PR 08-NOV-2000: 2000US-0244600
 PR 08-NOV-2000: 2000US-0244610
 PR 08-NOV-2000: 2000US-0244611
 PR 08-NOV-2000: 2000US-0244613
 PR 17-NOV-2000: 2000US-0249207
 PR 17-NOV-2000: 2000US-0249208
 PR 17-NOV-2000: 2000US-0249209
 PR 17-NOV-2000: 2000US-0249210
 PR 17-NOV-2000: 2000US-0249211
 PR 17-NOV-2000: 2000US-0249212
 PR 17-NOV-2000: 2000US-0249213
 PR 17-NOV-2000: 2000US-0249214
 PR 17-NOV-2000: 2000US-0249215
 PR 17-NOV-2000: 2000US-0249216
 PR 17-NOV-2000: 2000US-0249217
 PR 17-NOV-2000: 2000US-0249218
 PR 17-NOV-2000: 2000US-0249244
 PR 17-NOV-2000: 2000US-0249245
 PR 17-NOV-2000: 2000US-0249264
 PR 17-NOV-2000: 2000US-0249265
 PR 17-NOV-2000: 2000US-0249297
 PR 17-NOV-2000: 2000US-0249299
 PR 17-NOV-2000: 2000US-0249300
 PR 01-DEC-2000: 2000US-0250160
 PR 01-DEC-2000: 2000US-0250191
 PR 05-DEC-2000: 2000US-0251030
 PR 05-DEC-2000: 2000US-0251988
 PR 05-DEC-2000: 2000US-0256719
 PR 06-DEC-2000: 2000US-0251474
 PR 08-DEC-2000: 2000US-0251856
 PR 08-DEC-2000: 2000US-0251858
 PR 08-DEC-2000: 2000US-0251869
 PR 08-DEC-2000: 2000US-0251989
 PR 11-DEC-2000: 2000US-0251990
 PR 05-JAN-2001: 2001US-0253678
 PA (HUMA-) HUMAN GENOME SCI INC.
 XN Rosen CA, Harash SC, Ruben SM;
 XN WPI: 2001-488782/53.
 DR P-PSDB; AA021387.
 XN

New polynucleotides and polypeptides for diagnosing, treating, preventing or prognosis e.g. diseases of disorders of the nervous, musculoskeletal, excretory, gastrointestinal, reproductive, and respiratory systems

Claim 1: SEQ ID No 731: 642pp: English.

The invention relates to novel nucleic acids encoding novel human foetal antigens, the nucleic acids and proteins are used to prevent, treat (e.g. by gene therapy) or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the antigens can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotherapy. The polypeptides can also be used

CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. Numerous
 CC examples of diseases and disorders treated by the nucleic acids and
 CC proteins are given in the specification. The present sequence

alignment_scores.

Quality: 97.50 length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent identity: 66.667

alignment_block.

US-09-557-262-4_COPY_246_272 x AAS34207

Aliqu seq 1/1 to: AAS34207 from: 1 to: 503

1 PheTyrPheAsnLysIleLysGlyPheTyrCysLeuArg.....As 14
 |||||
 6 TCTATTTCACAAAATCAAGGAGTTCCTTGGCTAAAGCAAGGCAGAGA 55
 |||||
 14 pSerGlyArgAspAlaGlyCysLeuHisGluSerLysGlyArg 27
 |||||
 56 CAAAGTGTGCTTAAAGTATTAAGAAAGCAAGAGTGG 45

seq_name: /SIUS5/jcgdata/jcgenet4/genetseqp_emb1/NA1999.DAT:AA37250

seq_documentation_block:

LC AAX37250 standard, LNA, 3658 BP.

XX AC AAX37250;

XX XJ 20-JUL-1999 (first entry)

XX DE Human 3 OST-4 encoding DNA.

XX KW Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 KW heparan sulfate, thrombotic disorder; deep vein thrombosis;
 KW pulmonary embolism; coagulation enzyme inactivation; ss.

XX OS Homo sapiens.

XX PN WO9922005-A2.

XX PD 06-MAY-1999.

XX PE 23-OCT-1998; 98WO-0522597.

XX PK 31-OCT-1997; 97US-0065437.

XX PK 24-OCT-1997; 97US-0062762.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX PI Fritze LMS, Liu J, Rosenberg RD, Schwartz JJ, Shworak NW;

XX PI Zhang L;

XX WP: 1444-12469/26.

XX P-PSDB; AAY17067.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related

XX polynucleotides

PS Claim 5; Page 86-89; 95pp; English.

XX The invention relates to nucleic acid molecules (AAX37245-X37250)
 CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs).
 CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues
 CC within a preparation of glycosaminoglycan or proteoglycan
 CC polysaccharides. 3-OST-1 can be used for enriching the
 CC antithrombin binding fraction in a preparation of heparan sulfates (HS).
 CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 CC HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic
 CC function) can be used to determine partial sequence information for

OM of: US-09-557-262-4_copy_246_272.p2n.rge:Entl * out_fmt:msl fsl

Date: Jul 17, 2002 5:59 AM

About: Results were produced by the GeneCore software, version 4.5.

Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed_p2n.model -DEV=xlp
-G=/cdm2.1/us09/5572624/copy_246_272.p2n.rge -CAIOP=12.000
-DB=GenEmbl -QWTF=fastap -SUFFIX p2n.rge -CAIOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -DOPEXT=0.000 -DOPEXT=0.000
-QGAPEXT=4.500 -GAPEXT=0.050 -XGAPEXT=10.000 -XGAPEXT=0.500
-FGAPEXT=6.000 -GAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DEPEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=prf
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTWTF=prf
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US095572624_COPY_246_272 -NCP0=6 -ICFU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block.

Query: US-09-557-262-4_COPY_246_272

Query length: 27

Database: GenEmbl.*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 7102.980000

score_list:

Sequence	Strd	Orig	%Score	Len	Documentation
gb_pr:AF019386	155.00	422.73	3.5e-15	1305	AF019386 Homo sapiens heparan
gb_pr:AC073991	155.00	379.88	8.5e-13	159838	AC073991 Homo sapiens RA
gb_pr:AC006230	155.00	479.48	9e-13	166908	AC006230 Homo sapiens chrom
gb_to:AF177430	152.00	417.85	7.2e-15	916	AF177430 Paltus norvegicus 3
gb_to:BC009133	152.00	411.96	1.4e-14	1657	BC009133 Mus musculus, heparan
gb_to:AF019385	152.00	411.81	1.4e-14	1685	AF019385 Mus musculus, heparan
gb_btq:AC084416_3	152.00	374.56	1.7e-12	110000	Continuation (4 of 6) of AC0
gb_btq:AC098713	152.00	368.38	3.7e-12	220128	AC098713 Mus musculus chrom
gb_in:AC004657	114.00	268.31	1.4e-06	76618	AC004657 Drosophila melanoga
gb_btq:AC018336	114.00	265.95	2.1e-06	110482	AC018336 Drosophila melanoga
gb_in:AC003628	114.00	269.31	4.9e-06	242458	AC003628 Drosophila melanoga
gb_in:AF003378	114.00	257.38	5.7e-06	261090	AF003378 Drosophila melanoga
gb_pr:AF019378	97.50	252.37	1.1e-05	2211	AF019378 Homo sapiens heparan
gb_pr:AC093511	97.50	212.72	0.0017	189234	AC093511 Homo sapiens chrom
gb_btq:AC093359	97.50	211.43	0.0021	218502	AC093359 Mus musculus clone
gb_btq:AC021445	96.50	209.18	0.0027	204621	AC021445 Mus musculus clone
gb_pr:AF105377	88.50	227.20	0.0003	2532	AF105377 Homo sapiens heparan
gb_pr:AF105376	88.50	225.19	0.0004	2546	AF105376 Homo sapiens heparan
gb_pr:AK023723	88.50	224.31	0.0004	2408	AK023723 Homo sapiens cDNA Fl
gb_pr:AC005375	88.50	188.42	0.0369	149030	AC005375 Homo sapiens chrom
gb_pr:AC005224	88.50	187.92	0.0419	166687	AC005224 Homo sapiens chrom
gb_to:AF168992	87.50	226.09	0.0003	1656	AF168992 Mus musculus D-glyco
gb_pr:HS439A6	87.50	197.48	0.0124	41355	AL011723 Human DNA sequence
gb_pr:AC005163	87.50	192.14	0.0244	75108	AC005163 Homo sapiens chrom
gb_pr:AC006640	87.50	191.40	0.0268	81574	AC006640 Homo sapiens hpl3
gb_btq:AC012180	87.50	181.35	0.0974	320997	AC012180 Homo sapiens chrom
gb_btq:AL601489	87.50	179.22	0.1279	320996	AL601489 Mus musculus chrom
gb_pr:AF105374	86.50	221.72	0.0005	1968	AF105374 Homo sapiens heparan
gb_pr:AF105375	86.50	221.72	0.0005	1968	AF105375 Homo sapiens heparan
gb_btq:AC027102	86.50	182.95	0.0843	168612	AC027102 Homo sapiens chrom
gb_pr:HUAC002287	86.50	181.05	0.1011	188616	AC02287 Homo sapiens chrom
gb_btq:AC098715	83.50	169.93	0.4213	249269	AC098715 Mus musculus chrom
gb_pat:AX317978	75.50	189.39	0.0347	2115	AX317978 Sequence 18 from Pat
gb_pr:AL355498	75.50	150.53	5.07	165649	AL355498 Human DNA sequence
gb_btq:AC017305	71.50	143.52	12.46	96745	AC017305 Drosophila melanoga
gb_btq:AC068459	71.50	139.27	21.49	16366	AC068459 Homo sapiens clone
gb_btq:AC108928	71.50	139.24	21.79	16064	AC108928 Homo sapiens clone
gb_btq:AC073025	71.50	138.86	22.65	148247	AC073025 Homo sapiens chrom
gb_in:AC008337	71.50	138.78	22.87	169680	AC008337 Drosophila melanoga
gb_btq:AC095289	71.50	138.00	25.29	185314	AC095289 Paltus norvegicus c

gb_pr:AC005762 71.50 137.50 25.94 155337 1 AC005762 Homo sapiens chr
gb_in:AE003511 71.50 132.49 51.27 343807 1 AF003511 Drosophila melan
gb_in:AR037943 59.00 179.95 0.4246 22665 1 AR037943 Caenorhabditis elo
gb_in:AR037942 59.00 169.12 0.4673 2515 1 AR037942 Caenorhabditis elo
gb_in:AR037941 59.00 168.67 0.4948 2644 1 AR037941 Caenorhabditis elo

seq_name: gb_pr:AF019386

seq_documentation_block:

coding AF019386 1305 bp mRNA 1997 14 Nov 1997
DEFINITION Homo sapiens heparan sulfate 3-O-sulfotransferase 1 precursor
(3OST1) mRNA, complete cds.
ACCESSION AF019386
VERSION AF019386.1 GI:2618972
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1305)
AUTHORS Shworak, N.W., Liu, J., Fritz, L.M.S., Schwartz, J.J., Zhang, L.,
Logan, D., and Rosenberg, R.D.
TITLE Molecular cloning and expression of mouse and human cDNAs encoding
heparan sulfate D-glucosaminyl 3-O-sulfotransferase
JOURNAL J. Biol. Chem. 272 (44): 28068-28076 (1997)
MEDLINE 98010647
REFERENCE 2 (bases 1 to 1305)
AUTHORS Shworak, N.W., Liu, J., Fritz, L.M.S., Schwartz, J.J., Zhang, L.,
Logan, D., and Rosenberg, R.D.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
02139, USA
FEATURES
Location/Qualifiers
1..1305
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Isolated from lambda triplex brain cDNA library
from Clontech"
gene
1..1305
/gene="3OST1"
5'UTR
1..118
/gene="3OST1"
3'UTR
119..1278
/gene="3OST1"
CDS
119..1042
/gene="3OST1"
/function="rate limiting enzyme for synthesis of
anticoagulant heparan"
/note="heparan sulfate sulfotransferase; 3-OST-1;
interluminal GalII resident protein (retension mechanism
unknown)"
/codon_start=1
/product="heparan sulfate 3 O sulfotransferase-1
precursor"
/protein_id="AA884388.1"
/db_xref="GI:2618973"
/translation="MAALLISAVILVAGQELVPSDAPLGGGFLPKACTIQVNPVQ
VAPNSACQIPOTIIIGKQGTTHAIIEMI SIHPVAAAEVHFDFWPHYSGLQW
YLSQMPSPHGLIVETPKVPTSPKPVKYSMMPSRLIIILRDSRVISDYTOV
FYNMKPKIFPYSIEELVVRGPIVNVKALNPSLYHVMGNLFFELFHIVDQD
VLFFETFEIATKVEFLFKSTVNASNTYFNFTFYTLFESSECPQLHESKSHAPQ
VDFKLKLKHEVTEHFNKFFELVGRTEDMI"
119..1039
/gene="3OST1"
/product="heparan sulfate 3 O sulfotransferase-1"
260..268
/gene="3OST1"
/note="encodes potential N-linked site"
278..1039
/gene="3OST1"
/note="encodes presumptive sulfotransferase catalytic
domain"
527..535
misc_signal

```

/gene-"3OST1"
/Note-"encodes potential N-linked site: this site probably
not functional"
misc_signal
692..700
/gene-"3OST1"
/Note-"encodes potential N-linked site"
842..850
/gene-"3OST1"
/Note-"encodes potential N-linked site"
863..871
/gene-"3OST1"
/Note-"encodes potential N-linked site"
873..911
/Note-"encodes potential N-linked site"
873..911
/gene-"3OST1"
/Note-"encodes cysteine bridged peptide loop"
1043..1305
/gene-"3OST1"
1277..1283
/gene-"3OST1"
1305
/gene-"3OST1"
3'UTR
304 a 380 c 329 g 292 t
ORIGIN

alignment_scores:
  Quality: 155.00      length: 27
  Ratio: 5.741        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-4_COPY_246_272 x AF019386 ..
Align seg 1/1 to: AF019386 from: 1 to: 1305

1 PheTyRPhcAsnLysThrLysGlyPheTyRcysLeuArgAspSerGlyAr 17
|||||
854 TTTAT-TTTAA-AAAAT-AAAAT-TTTA-TAT-TATGAGAA-ATGAGAG-403

17 qASArqCysLeuHisGluSerLysGlyArg 27
|||||
904 GCAACGTGT-TTA-ATGAT-TC-AAAAGTCAG 934

seq_name: qb.pr:AC073991

seq_documentation_block:
LOCUS AC073991 159838 bp DNA linear PR: 07 NOV 2001
DEFINITION Homo sapiens BAC clone RP11-512120 from 4, complete sequence.
ACCESSION AC073991
VERSION AC073991.2 GI:14209802
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 159838)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res 8 (11): 1097-1108 (1998)
99063792
REFERENCE
2 (bases 1 to 159838)
Harris,A. and Kozlowicz,A.
The sequence of Homo sapiens BAC clone RP11-512120
Unpublished
JOURNAL
3 (bases 1 to 159838)
Waterston,R.H.
Direct Submission
AUTHORS
TITLE
JOURNAL
Submitted (08-Jul-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 159838)
Waterston,R.H.
Direct Submission
AUTHORS
TITLE

```

JOURNAL

Submitted (03-Jul-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 5 (bases 1 to 159838)
 Waterston,R.
 Direct Submission
 Submitted (07-Nov-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On May 26, 2001 this sequence version replaced gi:8980014.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Web site: http://genome.wustl.edu/qsc
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0512120

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone, and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The RPC111 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, H., Frenken, E., Tatenoe, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-669F3; the clone sequenced to the right is AC066230. Actual start of this clone is at base position 1 of RP11-512120; actual end is at base position 159838 of RP11-512120.

FEATURES

Location/Qualifiers	Source
1..159838	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="4"	
/map="4"	
/clone="RP11-512120"	
/clone_lib="RPC1-11"	
2..775	
/rpt_family="MERL_type"	
266..317	
/rpt_family="T-rich"	
770..792	
/rpt_family="AT-rich"	
845..2469	
/rpt_family="L1"	
2470..2766	
/rpt_family="Alu"	
2769..3291	
/rpt_family="MERL_type"	
3321..3441	

JOURNAL: Submitted (02 JUN 1999) Department of Genetics, Stanford Human Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA

COMMENT: On Jun 2, 1999 this sequence version replaced qi:4803914. Quality: Phrap Quality >40 99.5% of Sequence; Estimated Total Number of Errors is 0.3.

SIS Content:

SHGC-51595 G34132

WI-1308 G02734

SHGC-53120 G34506

SHGC-13460 G14876.

FEATURES

source

1..166998

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4p16"

/clone="C0287314"

/clone_lib="ROSWEI: PAPK CANCER PROJ - 11 Human Male PAC Library"

BASE COUNT 49526 a 33832 c 34587 g 49053 t

ORIGIN

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_246_272 x AC006230/rev ..

Align seq 1/1 to: reverse of: AC006230 from: 1 to: 166998

1 PheTyRPhcAsnLysThrLysGlyPheTyRcysLeuArgAspSerGlyAr 17

|||||

101629 TTCTACTTTACAAACCAAGCGCTTTATCGTGTGGGACACACCGCGG 101580

17 gAspArgcysLeuHisGluSerLysGlyArg 27

|||||

101579 GGACGGCTGTTCATGATGATGCAAGGCGCG 101549

seq_name: qb_ro:AF177430

seq_documentation_block:

LOCUS AF177430 936 bp mRNA linear pop 10-MAY-2001

DEFINITION Rattus norvegicus 3-O-sulfotransferase mRNA, complete cds.

ACCESSION AF177430

VERSION AF177430.1 GI:9957243

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 936)

AUTHORS Li,Z.Y., Hirayoshi,K. and Suzuki,Y.

TITLE Expression of N-deacetylase/sulfotransferase and 3-O-sulfotransferase in rat alveolar type II cells

JOURNAL Am. J. Physiol. 275 (2), L592-L595 (1998)

REFERENCE 2 (bases 1 to 936)

AUTHORS Li,Z.Y., Hirayoshi,K. and Suzuki,Y.

TITLE Direct Submission

JOURNAL Submitted (13-AUG-1999) Department of Ultrastructural Research, Institute for Frontier Medical Sciences, 53 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan

FEATURES

source

1..936

/organism="Rattus norvegicus"

/strain="Wistar"

/db_xref="taxon:10116"

/cell_type="type II epithelium"

1..936

/note="3-OST"

CDS

codon_start=1

/product="3-O-sulfotransferase"

/protein_id="AAC09284.1"

/db_xref="GI:9957244"

/translation="MTTITLIGAVLIIVAGPQIVPSIPAAAPDRIKQGLIPKVIILPEDTGGAAINGSTQQLPQTIIIGVPRKGTALLLEMLSLHVDVAAAEVHFPIWERYSUGLGWYLITOMPSSPHOLVETKIPAYETSPKVERIHSMNPTIRLLILLIRDSRVLSDYTQVLYNHILOKHKYPPIFDLLMRICRLNVQYKALNSLYHAHMLNMLRPPDGLCHITIVDRTFTIRPPEIKTSPETINASNFENKTFGFCVLCISGSDRCLUESKGR

BASE COUNT 224 a 292 c 224 g 196 t

ORIGIN

alignment_scores:

Quality: 152.00 Length: 27

Ratio: 5.640 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_COPY_246_272 x AF177430 ..

Align seq 1/1 to: AF177430 from: 1 to: 936

1 PheTyRPhcAsnLysThrLysGlyPheTyRcysLeuArgAspSerGlyAr 17

|||||

748 TTCTACTTTTAAATAAAATAAGGATTTATGCTGTGGGACACACCGCG 797

17 gAspArgcysLeuHisGluSerLysGlyArg 27

|||||

798 GCACCGCTGTTCATGATGATGCAAGGCGCG 828

seq_name: qb_ro:BC009133

seq_documentation_block:

LOCUS BC009133 1657 bp mRNA linear REF 12-JUL-2001

DEFINITION Mus musculus, heparan sulfate (glucosamine) 3-O-sulfotransferase 1, clone MGC:11450 IMAGE:3155049, mRNA, complete cds.

ACCESSION BC009133

VERSION BC009133.1 GI:1414662

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1657)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892 2590, USA

FEATURES

source

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue procurement: Iohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: http://www.hgsc.bcm.tmc.edu/cdna/

Contact: villalona@bcm.tmc.edu.

Villalona, D.K., Luna, P.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny,D.M., Gibbs,R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Classification at: http://image.llnl.gov

Series: IRAP Plate: 16 Row: 0 Column: 17

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA qi: 2618970.

Location/Qualifiers

source

1. .1657
 /organism="Mus musculus"
 /db_xref="locus:15476"
 /db_xref="taxon:10090"
 /clone "MCC:11450 IMAGE:3155049"
 /tissue_type="Mammary tumor, Brca1-/fl; MMTV-Cre model. 10 months old, gross tissue."
 /clone_lib="NCI_CGAP_Mam3"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

CDS

284..1219
 /codon_start=1
 /product="heparan sulfate (glucosamine)
 3-O-sulfotransferase 1"
 /protein_id="AAH09133.1"
 /db_xref="GI:14318663"
 /translation="MTLLLGAVLLVAQPOLVHSHPAAPGCLKQELLKPKVLIILPDD
 TGEIASNGSTQQLPWIIIGYRKGGTRALLEMLSLHGVAAAAENEVHFFWEHHYSQ
 GLWYLQMPFSSPHQIVVKETPAYFISPKVPIHSMPTIRLLILLRDPSERVLSQ
 YTOVLNHLQKHKPPYPIEDLLMRDGRNLKALNRSLYHAHMLNLRFPFLGHII
 VGDPLIRDPFPIQKVPFELKSPQINASNFYFNKTKGFCYCLPNSCKDRCIHFSGKR
 AHPQVDPKLLDKLHEYFHEPNKKEFLVGRTFDWH"
 BASE COUNT 398 a 439 c 393 g 427 t
 ORIGIN

alignment_scores

Quality: 152.00 length: 27
 Ratio: 5.630 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block

US-09-557-262-4_COPY_246_272 x MUS09133 ..
 Align seg 1/1 to: MUS09133 from: 1 to: 1657
 1 PheTyrPheAsnIysThrIysGlyPheTyrCysLeuArgAspSerGly;Ar 17
 1031 TTCTACTTTAAACAAACCAAGGGCTTCTACTCCCGGGACAGTGGCAA 1080

17 gSPArGcysLeuHisGluSerIysGlyArg 27

1081 GAGACGCTGTTTAAGAGTGAATGAAAGACGG 1111

seq_name: gb_ro:AF019385

seq_documentation_block:
 LOCUS AF019385 1685 bp mRNA linear 14-NOV-1997
 DEFINITION Mus musculus heparan sulfate D-glucosaminyl 3-O-sulfotransferase-1 precursor (30ST1) mRNA, alternatively spliced, complete cds.

ACCESSION AF019385

VERSION AF019385.1 GI:2618970

KEYWORDS

SOURCE

ORGANISM

Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 1685)
 Shworak, N. W., Liu, J., Fritze, I. M. S., Schwartz, J. J., Zhang, L.,
 Loebart, D., and Rosenberg, R.D.
 Molecular cloning and expression of mouse and human cDNAs encoding
 heparan sulfate D-glucosaminyl 3-O-sulfotransferase
 J. Biol. Chem. 272 (44), 28068-28079 (1997)

JOURNAL 98010647

MEDLINE 2 (bases 1 to 1685)

Shworak, N. W., Liu, J., Fritze, I. M. S., Schwartz, J. J., Zhang, L.,
 Loebart, D., and Rosenberg, R.D.

AUTHORS

TITLE Direct Submission

Submitted (14 AUG 1997) Biology, MIT, 31 Ames Street, Cambridge, MA

JOURNAL 02139, USA

FEATURES location/Qualifiers

1..1685

/organism="Mus musculus"

/strain="C3H/An"
 /db_xref="taxon:10090"
 /feature="cDNA:1..1685" note="cDNA clone: 1A (described in Shworak et al. 1996, J. Biol. Chem. 271:27071-82)"
 /note="poly(A)+ RNA isolated from postconfluent cultures"
 1..322
 /gene="30st1"
 /note="alternatively spliced; this version represents a Class 1 mRNA"
 1..1685
 /gene="30st1"
 216..217
 /gene="30st1"
 /note="splice variant site; many additional 3'OST-1 cDNAs (Class 2 to Class 9) contain extra 5' UTR exons between these nucleotides (extra 86 to 1629 bp)"
 323..382
 /gene="30st1"
 323..1258
 /gene="30st1"
 /function="rate limiting enzyme for the synthesis of anticoagulant heparan sulfate"
 /note="heparan sulfate sulfotransferase; 3-OST-1; interluminal Golgi resident protein (retension mechanism unknown)"
 /codon_start=1
 /product="heparan sulfate D-glucosaminyl 3-O-sulfotransferase 1 precursor"
 /protein_id="AAH84387.1"
 /db_xref="GI:2618971"
 /translation="MTLLLGAVLLVAQPOLVHSHPAAPGCLKQELLKPKVLIILPDD
 IGEIASNGSTQQLPWIIIGYRKGGTRALLEMLSLHGVAAAAENEVHFFWEHHYSQ
 GLWYLQMPFSSPHQIVVKETPAYFISPKVPIHSMPTIRLLILLRDPSERVLSQ
 YTOVLNHLQKHKPPYPIEDLLMRDGRNLKALNRSLYHAHMLNLRFPFLGHII
 VGDPLIRDPFPIQKVPFELKSPQINASNFYFNKTKGFCYCLPNSCKDRCIHFSGKR
 AHPQVDPKLLDKLHEYFHEPNKKEFLVGRTFDWH"
 383..1255
 /gene="30st1"
 /note="confirmed by N-terminal analysis"
 /evidence="experimental"
 /product="heparan sulfate D-glucosaminyl 3-O-sulfotransferase-1"
 476..484
 /gene="30st1"
 /note="encodes N-linked glycosylation sites"
 /evidence="experimental"
 494..1255
 /gene="30st1"
 /note="encodes presumptive sulfotransferase catalytic domain"
 908..916
 /gene="30st1"
 /note="encodes N-linked glycosylation sites"
 /evidence="experimental"
 1058..1066
 /gene="30st1"
 /note="encodes N-linked glycosylation sites"
 /evidence="experimental"
 1079..1087
 /gene="30st1"
 /note="encodes N-linked glycosylation sites"
 /evidence="experimental"
 1100..1129
 /gene="30st1"
 /note="encodes cystine bridged peptide loop"
 /evidence="experimental"
 1259..1285
 /gene="30st1"
 1532..1537
 /gene="30st1"
 1556
 /gene="30st1"
 1586..1591
 /gene="30st1"

[illegible]

Quality levels above 40 are expected to have less than 1 error in 10,000 bp
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
 source 1..189234
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="CID-2507C6"
 BASE COUNT 55941 a 38428 c 38216 g 56658 t 1 others
 ORIGIN

alignment_scores:
 Quality: 97.50 Length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 66.667

alignment_block:
 US-09-557-262-4_copy_246_272 x AC093511/rev
 Align seg 1/1 to reverse of: AC093511 from 1 to 189234

1 PheTyPheAsnLysThrLysGlyPheLysCysLeuArg.....AS 14
 114117 TCTATTTCACAAACCAAGGAGCTTTCTTTTAAAGAAAGAAAGA 114068
 14 pSerGlyArgAspArgCysLeuHisGluSerLysGlyArg 27
 114067 CAGCAGTCCCGAGGCTCTTAGGCACAGCAAGGTCGG 114028

seq_name: qb_hlg.AC093359

seq_documentation_block.
 LOCUS AC093359 218502 bp DNA linear HPC 20-AUG-2001
 DEFINITION Mus musculus clone RP23-57A19, WORKING DRAFT SEQUENCE, 16 unordered
 pieres

ACCESSION AC093359
 VERSION AC093359.1 GI:15213904
 KEYWORDS HPG: HTGS_PHASE1: HTGS_DPRAFT: HTGS_FULLTOP
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
 Brown,A., Camarata,J., Campione,A., Chang,J., Chazaro,B.,
 Choquet,Y., Collins,S., Collins,S., Collins,S., Cook,A.,
 Cooke,P., DeArnell,K., Dewar,K., Bied,J.S., Bied,J.S., Bied,J.S.,
 Ferrel,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glode,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Haas,B., Heaford,A., Horton,L., Hulme,W., Iliiev,I., Johnson,R.,
 Jones,C., Kamal,A., Karalas,A., Kells,C., Larocque,K.,
 Lamazares,R., Landers,T., Lecheczy,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

TITLE
 JOURNAL
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

Center project name: 114290
 Center clone name: 57_A_19

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731

Consensus quality: 21338 bases at least Q40
 Consensus quality: 21548 bases at least Q40
 Consensus quality: 21643 bases at least Q20

Insert size: 235000; agarose-lp
 Insert size: 217002; sum of contigs

Quality coverage: 8.7 in Q20 bases; 9.4 in Q20 bases; sum of contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 16 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 59151: contig of 59151 bp in length
 59152 59251: gap of 100 bp
 59252 6031: contig of 1860 bp in length
 6032 6041: gap of 100 bp
 6042 6206: contig of 162 bp in length
 6207 6216: gap of 100 bp
 6217 6411: contig of 1948 bp in length
 6412 6421: gap of 100 bp
 6422 6579: contig of 1582 bp in length
 6580 6589: gap of 100 bp
 6590 6908: contig of 3145 bp in length
 6909 6938: gap of 100 bp
 6939 74310: contig of 5172 bp in length
 74311 74410: gap of 100 bp
 74411 82566: contig of 8156 bp in length
 82567 82666: gap of 100 bp
 82667 93394: contig of 10728 bp in length
 93395 93494: gap of 100 bp
 93495 105599: contig of 12105 bp in length
 105600 105699: gap of 100 bp
 105700 117559: contig of 11860 bp in length
 117560 117559: gap of 100 bp
 117600 132463: contig of 19804 bp in length
 132464 132563: gap of 100 bp
 132564 157070: contig of 19507 bp in length
 157071 157170: gap of 100 bp
 157171 177936: contig of 20766 bp in length
 177937 178036: gap of 100 bp

* 178037 205897: contig of 27861 bp in length
 * 205898 205997: gap of 100 bp
 * 205998 218502: contig of 12505 bp in length.

FEATURES

source

Location/Qualifiers
 1..218502

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-57A19"
 /clone_lib="RPCL-23 Female Mouse BAC"

1..59151

/note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

misc_feature

59252..60311

/note="assembly_fragment"

misc_feature

60412..62063

/note="assembly_fragment"

misc_feature

62164..64111

/note="assembly_fragment"

misc_feature

64212..65793

/note="assembly_fragment"

misc_feature

65894..69038

/note="assembly_fragment"

misc_feature

69139..774310

/note="assembly_fragment"

misc_feature

74411..82566

/note="assembly_fragment"

misc_feature

82667..93394

/note="assembly_fragment"

misc_feature

93495..1105599

/note="assembly_fragment"

misc_feature

105700..117559

/note="assembly_fragment"

misc_feature

117660..137463

/note="assembly_fragment"

misc_feature

137564..157070

/note="assembly_fragment"

misc_feature

157171..177936

/note="assembly_fragment"

misc_feature

178037..205897

/note="assembly_fragment"

misc_feature

205998..218502

/note="assembly_fragment"

misc_feature

clone_end:17

vector_side:right"
 BASE COUNT 64182 a 43459 c 44675 g 64786 t 1500 others
 ORIGIN

alignment_scores:

Quality	97.50	Length	30
Ratio	4.239	Gaps	1
Percent Similarity	76.667	Percent Identity	66.667

alignment_block:

US-09-557-262-4_COPY_246_272 x AC093359 ..

Align seq 1/1 to: AC093359 from: 1 to: 218502

```

1  phtyrlpAsnTyrLysGlyPheTyrCysLeuArg.....As 14
|||||
35843  TTTTACTTCACAAACCAAGCGTTCCTCCCTCCCTGAGAGAGCCACAGAA 35892
|||||
14  pSerGlyArqAspArqCysLeuHisGluSerLysGlyArq 27
|||||
35893  CAGTACGCGCCCAAGATGCTTGTGGCAAGCAAGCAAGGTCGG 35932
|||||

```


Date: Jul 17, 2002 6:03 AM

Copyright (c) 1993-2000 CompuGen Ltd.

```
-MODEL=frame; b2u_model -DEV-x'1'
```

-DB-N_Geneseq_032802 -QPMT-fastaf -SUPPLX-p23 ref -CAP00-

Query: US-09-557-262-4 COPY 21 303

Database: N Geneseq 032802.*

Database sequences: 1730430
Database length: 858457221

000009 : ZL : () 3177 NO 1600

References

/s/055/gcga

7SID55/9299

5013/STDS/5013/

06/07/2011

/sids5/qcqd.

pub/yulis/

psib/gcgd

/s:1b55/qcqd/

/SIDS5/qcqd

/S10S5/gcqd.

/sids5/qcqd

/SID\$5/arcqd

[/SID\\$5/qcqd](#)

/STDS5/acad

/STD55/qcga
/STD55/qcga

/sips/gcga
/sips/gcga

/STDS/gcga
/STDS/gcga

pbcb/sgts/

/51055/gcgg

/51155/qcqd

pubs/55DIS/

alignment_block:

US-09-557-262-4_copy_21_303 x AAZ37246

Align seq 1/1 to: AAZ37246 from: 1 to: 1305

```

1  ArgProAlaGluGlyGlyGlnGlnGluLeuLeuArgGlySerAlaGlnGlnL 17
|||||
179 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 228
|||||
17  uGluAspAspValArgAspGlyValAlaProAsnGlySerAlaGlnGlnL 34
|||||
229 CACATATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
|||||
34  euProGlnThrIleIleIleGlyValArgGlyGlyGlyThrArgAlaLeu 50
|||||
279 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 328
|||||
51  LeuGluMetLeuSerLeuHisProAspValAlaAlaAlaGluAsnGluVa 67
|||||
329 CCGGACATGCTTACGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 378
|||||
67  IHisPhePheAspTrpGluGlnHisTyrSerHisGlyLeuGlyTrpTyrL 84
|||||
379 CCACCTCTTCACCTCGGAGGAGATATACAGCCACCGGCTTGGGCTGGTACC 428
|||||
84  euSerGlnMetProPheSerTrpProHisGlnLeuThrValGlnGlyThr 100
|||||
429 TCAACACATGCTTCTTCTGAGGAAATATATATATATATATATATATAT 478
|||||
101 ProAlaTyrPheThrSorProIysValProGluArgValTyrSerMetAs 117
|||||
479 CCGCGGTAATTCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528
|||||
117 nProSerIleArgLeuLeuLeuLeuArgAspProSerGluArgValL 134
|||||
529 CCGGGLCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 578
|||||
134 euSorAspTyrThrGlnValIlePheTyrAsnHisMetGlnIlyHisIysPro 150
|||||
579 TATCTGATATATATATATATATATATATATATATATATATATATATAT 628
|||||
151 TyrProSerIleGluGluPheLeuValArgAspGlyValArgAsnValAs 167
|||||
629 TACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 678
|||||
167 pTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetGlnAsnTrpL 184
|||||
679 CTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
|||||
184 euArgPhePheProLeuArgHisIleHisIleValAspGlyAspArgLeu 200
|||||
729 TGGCTTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778
|||||
201 IleArgAspTrpPheTrpGluGlnGlyValGluArgPheLeuLysLe 217
|||||
779 ATCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
|||||
217 uSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysThrLysGlyP 234
|||||
829 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 878
|||||
234 heTyrCysLeuArgAspSerGlyValArgAspArgCysLeuHisGluSer 250
|||||
879 TTTTACTGTTTATGAGAAATAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928
|||||
251 GlyArgAlaHisProGlnValAspPheTrpLysLeuAsnLysLeuHisG 267
|||||
929 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 978
|||||
267 uTyrPheHisSerProAsnAspValTyrPheThrGlnLeuValGlyArgThr 283
|||||
979 ATATTTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027

```

seq_name: /SID55/4-3-2002/4-3-2002-Seq/4-3-2002-Seq-ent/NA2030 DAT AAZ36687

seq_documentation_block:

AAZ36687 standard; DNA; 1305 BP.

XX AAZ36687;

XX 13 MAR-2000 (first entry)

XX Nucleotide sequence for GenBank accession number AF019386.

XX Stimulus-regulated nucleic acid, sequence profile, nucleic acid level;
 XX differentially expressed nucleic acid, disease state; cancer;
 KW autoimmune disease; infectious disease; aging; developmental disorder;
 KW proliferative disorder; neurological disorder; toxicity;
 KW treatment resistance; differential expression; drug discovery;
 KW growth factor; epidermal growth factor; radiation; stress; pathogen; ss.

XX Homo sapiens.

XX W09955913-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99W0-US09119.

XX 27-APR-1998; 98US-0083331.

XX 27-AUG-1998; 98US-0098070.

XX 04-FEB-1999; 99US-0118624.

XX (KIMM-) KIMMEL CANCER CENT SIDNEY.

XX McClelland M, Welsh J, Trenkle T;

XX WPI: 2000-086388/07.

XX Measuring expression of low abundance reduced complexity target nucleic

XX acid molecules

XX Disclosure, Fig 14, 187pp, English.

XX AAZ36687 236725 represent nucleotide sequences from stimulus-regulated
 CC nucleic acid molecules. The sequences represent a profile of sequences
 CC which can function as targets in the method of the invention. The
 CC specification describes a method for measuring the level of two or more
 CC nucleic acid molecules in a target. The method comprises contacting a
 CC probe with an arbitrarily or statistically sampled target and detecting
 CC the amount of specific binding of the target to the probe. The methods
 CC can be used to identify differentially expressed nucleic acid molecules
 CC associated with disease states, such as cancer, autoimmune disease,
 CC infectious disease, aging, developmental disorder, proliferative
 CC disorder or neurological disorder. Alternatively the methods can be
 CC used to assess the efficacy or toxicity of or a resistance to a
 CC treatment. Also the methods can be used to determine differential
 CC expression of nucleic acid molecules in response to a stimulus, e.g. a
 CC chemical, drug or growth factor (especially epidermal growth factor),
 CC radiation, stress or a pathogen. The methods can also be used to
 CC determine co-regulated genes that can be potential targets for drug
 CC discovery.

XX Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:

Quality: 1526.00 Length: 284
 Ratio: 5.392 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US 09-557-262-4_COPY_21_303 x AAZ36687

Align seq 1/1 to: AAZ36687 from: 1 to: 1305

heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant; saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST; heparan sulfate, thrombotic disorder, deep vein thrombosis; pulmonary embolism; coagulation enzyme inactivation; ss.

Mus musculus.

W09922005-A2.

06-MAY-1999.

23-OCT-1998; 98WQ-US22597.

31-OCT-1997: 97US-0065437.

24-OCT-1997; 17661-1DO-47 97US-0062762.

(MASI) MASSACHUSETTS INST TECHNOLOGY.

Eritze LMS, Liu J, Rosenberg PD, Schwartz JT, Shwartzak NW, Zhang L;

WPI; 1999-312968/26.

p-PSDB, AAY17062.

Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related polynucleotides

Claim 5; Page 71-72; 95pp; English.

The invention relates to nucleic acid molecules (AA37245-X37250) encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs). The 3-OST proteins can be used for 3-O-sulfating saccharide residues within a preparation of glycosaminoglycan or proteoglycan polysaccharides. 3-OST-1 can be used for enriching the antithrombin-binding fraction in a preparation of heparan sulfates (HS). 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic function) can be used to determine partial sequence information for complex polysaccharides. The 3-OST proteins, genes and antibodies are also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs are especially used to generate anticoagulant pentasaccharides, which may be used to treat thrombotic disorders such as deep vein thrombosis and pulmonary embolism. *Regulation of enzyme inactivation by antithrombin is enhanced by complexing of antithrombin with endothelial cell surface HS proteoglycans.* This is responsible for the non thrombotic properties of blood vessels. The present sequence represents a mouse 3-OST-1 encoding DNA.

Sequence 1685 BP; 422 A; 426 C; 382 G; 455 T; 0 other.

ignment scores:

Ratio:	1343.00	length:	279
Ratio:	4.993	Gaps:	0
Ratio:	96.416	Percent Identity:	88.883

alignement_block: US-09-557-262-4 COPY 21 303 X AAX37245

Align seq 1/1 to: AAX37245 from: 1 to: 1685

5 LeucylcyslnGlnIleLeuLeuArgLysAlaThrLeuGlnASPspva 21
||| ||||| ||||||| ||||| |||
407 CTCAACAGCAGCACGCCCTTCAGCAAGTGCATATTCTCCACAGGACAC 456

21 l a r q A s p G l y V a l A l a P r o C a s n G l y S e r A l a G l u G l i n I e u P r o G l u T h r I 38
 :
457 C G G A G A A G G C A A G A G A T C T A A T G T T C A C A N A G N Y C C G C A A G A C A A 506

38 | e l l e l l e G l y V a l A t g L y s G l y G l y T h r A r g A l a I e u l e u G l u M e t I e u 54
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
507 T C A T C A T T G G G T G G C A A G G T A T A C C A G C C T T A G A C A T G C T C 556

```

55 SerLeuHisProAspValAlaAlaGluAsnGluValIleHisPhePheAs 71
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
557 AGCCGTGAAGCCTGGTGGTCACATCAAAAGAACAGTGCAATTCTTTCA 606
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 pTPGluGluHisIstYrSerHisGlyLeuClYTrpIyrIeuSerGlnMetP 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
607 CTAAATAAATAATTAATATATATATATATATATATATATATATATAT 656
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 ropHeSerTripproHisGlnIeuthrValGluIylsThrProAlatyrphe 104
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 CCTTTCCTCTGCCCTTACACAGCTCACCGGTGCAGACACACACCCTATTC 706
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 ThrSerProLysValProGluArqValItyrSerMetAsnProSerIleAr 121
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
707 ACTTCGCCCAAGTGCCTTCAGAGAATCCACAACAAGAACCCACCATGG 756
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 gLeuleuleutIleuArqAspProSerGluArqValIleuSorAsplyrr 138
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
757 CCTGCTACTTATATCTGATATATATATATATATATATATATATATAT 806
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
138 hrGlnValPheTyrrAsnHisMetGlnLysHisLysPtyrProSerIle 154
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
807 CCCAGATGTGTAATAACCACTTCAGAAAGCAACAACCTATGCACCAAT 856
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 GlcGluPheLeuValArqAspGlyArgLeuAsnValasfrrLySalale 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
857 GAGCACTCTCTAAAGCGGATCGGCTGGAGCTGGCACTATCAAGGCTCT 906
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 uAsnArsSerIeutyrrHisValHisMetGlnAsnTrpleuArgPhePhoP 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
907 TAAAGATATATATATATATATATATATATATATATATATATATAT 956
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
188 rOleuArqHisIstHisIleValAspGlyAsnArqLeuIcarGAspbro 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
957 CGTGTGGCTCAATATATATATATATATATATATATATATATATAT 1006
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 PheProglLalleGluLysValGluArqPheLeuLysLeuSerProGluLi 221
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1007 TATATACAGATCTCAAGATCTGAAAAGATATATATATATATATAT 1056
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 oASnAlisScrAsnphetyrrPheAsnLysThrlySGlyPheTyrrCSySLCA 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1057 CAAGCCTCTGAACTCTACTTTATCAAAAACCAAGGGCTTCATCTGGCTGC 1106
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 rGAspserCItyArqASPArqCysIeuHisGIssetrCysSCLYArgALahis 254
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1107 GGGATACITGGCAAGCACTGCTGCTTACATCAAGTCCAAAGGCGCGGCAAC 1156
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
255 ProGluValAspfLoLysLeuLeuAsnLysleuHISGduLyrrPheHisLi 271
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1157 CCCCAAGTGGATGCCAACTACTTCATAAATGCCACGAATATCTCAICA 1206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
271 uproAsniYstIystPhebeGluIeuValICIYargthr 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1207 GCATATATATATATATATATATATATATATATATATATATATAT 1243
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name = /SIN85/gcgdata/geneseq/genomeseq.tbl/NA2001A.baf.AHA4490B
seq_documentation_block:
ID ID ABA44908 standard; DNA; 452 bp.
XX XX
AC ABA44908;
XX XX
XX XX 01-FEB-2002 (first entry)
XX XX Human breast cell single exon nucleic acid probe #3603.
KW KW Human; microarray; single exon probe, gene expression; breast;
KW KW disease; cancer; ss.
OS OS Homo sapiens.
XX XX W0200157271-A2.

```


alignment_scores: Quality: 829.00 Length: 150
 Ratio: 5.527 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_21_303 x ABA55377/rev ..

Align seg 1/1 to reverse of: ABA55377 from: 1 to: 452

65 AsnGluValHisPhePheAspTyrClnGluHisTyrSerHisGlyLeuG1 81
 |||||
 451 AACGAGGTCACCTTTCTGACTGGAGAGCATTAACAGCAGGCTTGGG 402
 81 yTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
 |||||
 401 CTGGTACCTCAAGCAAGATGCTTTCTGCTGCTGCTGCTGCTGCTGCT 452
 98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
 |||||
 351 AGAAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 402
 115 SerMetAsnProSerTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 131
 |||||
 301 AATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452
 131 uArgValLeuSerAspTyrThrGluValPheTyrAsnHisMetGlnLysH 148
 |||||
 251 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
 148 IsLysProTyrProSerTyrLeuValLeuValLeuValLeuValLeuVal 164
 |||||
 201 ACAAGCCCTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452
 165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
 |||||
 151 AATGTGATCAATCAAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
 181 nAsnTrpLeuArgPhePheProLeuArgHisIleHisIleValAspGlyA 198
 |||||
 101 GAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452
 198 spArgLeuIleArgAspProPheProGluLeuLeuLysValHisuArgPhe 214
 |||||
 51 ACCGCTCTCATCAGGACCCCTTCCCTGAGATCCAAAGCTCGAGAGGTT 2

seq_name: /Snp5/ycgdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA25103

seq_documentation_block:

ID ABA25103 standard; DNA; 452 BP.

XX AC ABA25103;

XX DT 23-JAN-2002 (first entry)

XX DE Probe #356q for gene expression analysis in human heart cell sample

XX KW Human; gene expression; heart; microarray; vascular system; probe;

XX KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX KW congenital heart disease; ss.

XX OS Homo sapiens.

XX PN WC200157274-A2.

XX XX 09-AUG-2001.

XX XX 30-JAN-2001; 2001W0-NS00666.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632456.

PP 21-SEP-2000; 2000US-0234587.
 PR 27-SEP-2000; 2000US-0234589.
 XX 04-OCT-2000; 2000US-0234593.
 XX (MORF-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Harzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

XX claim 1: SEQ ID No 356q; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:

Quality: 829.00 Length: 150
 Ratio: 5.527 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_21_303 x ABA25103/rev ..

Align seg 1/1 to reverse of: ABA25103 from: 1 to: 452

65 AsnGluValHisPhePheAspTyrClnGluHisTyrSerHisGlyLeuG1 81
 |||||
 451 AACGAGGTCACCTTTCTGACTGGAGAGCATTAACAGCAGGCTTGGG 402
 81 yTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
 |||||
 401 CTGGTACCTCAAGCAAGATGCTTTCTGCTGCTGCTGCTGCTGCTGCT 352
 98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
 |||||
 351 AGAAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
 115 SerMetAsnProSerTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 131
 |||||
 301 AATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 252
 131 uArgValLeuSerAspTyrThrGluValPheTyrAsnHisMetGlnLysH 148
 |||||
 251 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
 148 IsLysProTyrProSerTyrLeuValLeuValLeuValLeuValLeuVal 164
 |||||
 201 ACAAGCCCTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 152
 165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
 |||||
 151 AATGTGATCAATCAAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 102
 181 nAsnTrpLeuArgPhePheProLeuArgHisIleHisIleValAspGlyA 198
 |||||
 101 GAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 52

CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention
 XX
 SQ Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:
 Quality: 829.00 length: 150
 Ratio: 5.527 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-557-262-4_copy_21_303 x AAK29079/rev ..

Align seq 1/1 to reverse of: AAK29079 from: 1 to: 452

```

65 AsnGluValHisPheAspTrpGluHisTyrSerHisGlyLeuG1 81
|||||
451 AAGCAGTCTCAGTCTCTCCAGCTGGAGAGAGATTAAGCAGAGGTTGGG 402

81 YTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
|||||
401 CTGATGCTCAGAGATGAGATGATTTCTGTGGGAAATTAATTAATTA 352

98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
|||||
351 AGAAGACCCCGCGTATTTTCAGCTCCGCTCCCAAGCTGCTTCACGCTCA 302

115 SerMetAsnProSerIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 131
|||||
301 AGCATGAACCCGCTCCATCCGCTCTCTCCGCTCAATCCCTCCGAGACCT 252

131 uArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMetGlnLysH 148
|||||
251 GCGATGCTATTTATTTAAATTAATTAATTAATTAATTAATTAATTA 202

148 isLysProTyrProSerIleGluGluPheLeuValArgAspGlyArgLeu 164
|||||
201 ACAAGCCCTACCGCTCCATCCGAGGAGTTTCCTGCTCCGCGATCCGAG 152

165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
|||||
151 AATGTCGATTAAGCGCTCTTAAGCGCTCTTAAGCGCTCTTAAGCGCT 102

181 nAsnTrpLeuArgPheProLeuArgHisIleHisIleValAspGlyA 198
|||||
101 GAACGTGCTGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 52

198 spArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 214
|||||
51 ACCGCTCTATCAGGAGCCCTCTCTCTGAGATCCAAAGGCTGAGAGCTTC 2

```

seq_name: /SID55/gcdata/geneseq/geneseq-emb1/NA2001A.DAT.AAI35025

seq_documentation_block:

ID AAI35025 standard; DNA: 452 BP.

AC AAI35025,

DT 17-OCT-2001 (first entry)

DE Probe #3711 used to measure gene expression in human placenta sample.

XX Probe: microarray; human: placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

OS

XX WO200157272-A2.

PN

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-0500663.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-APR-2000; 2000US-0642666.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0649359.
 XX 04-OCT-2000; 2000GB-004463.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG. Hanzel DK. Chen W. Pank DP.
 XX WPI: 2001-488897/53.
 XX
 XX Human genome derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX Claim 25; SEQ ID NO 3711; 654pp; English.
 XX
 XX The present invention relates to single exon nucleic acid probes (SNP)
 XX The present sequence is one such probe. The probes are useful for
 XX producing a microarray for predicting, measuring and displaying gene
 XX expression in samples derived from human placenta. The probes are useful
 XX for antenatal diagnosis of human genetic disorders.
 XX
 XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:

Quality: 829.00 length: 150
 Ratio: 5.527 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-557-262-4_copy_21_303 x AAI35025/rev ..

Align seq 1/1 to reverse of: AAI35025 from: 1 to: 452

```

65 AsnGluValHisPheAspTrpGluHisTyrSerHisGlyLeuG1 81
|||||
451 AAGCAGTCTCAGTCTCTCCAGCTGGAGAGATTAAGCAGAGGTTGGG 402

81 YTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
|||||
401 CTGATGCTCAGAGATGAGATGATTTCTGTGGGAAATTAATTAATTA 352

98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
|||||
351 AGAAGACCCCGCGTATTTTCAGCTCCGCTCCCAAGCTGCTTCACGCTCA 302

115 SerMetAsnProSerIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 131
|||||
301 AGCATGAACCCGCTCCATCCGCTCTCTCCGCTCAATCCCTCCGAGACCT 252

131 uArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMetGlnLysH 148
|||||
251 GCGATGCTATTTATTTAAATTAATTAATTAATTAATTAATTAATTA 202

148 isLysProTyrProSerIleGluGluPheLeuValArgAspGlyArgLeu 164
|||||
201 ACAAGCCCTACCGCTCCATCCGAGGAGTTTCCTGCTCCGCGATCCGAG 152

165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
|||||
151 AATGTCGATTAAGCGCTCTTAAGCGCTCTTAAGCGCTCTTAAGCGCT 102

181 nAsnTrpLeuArgPheProLeuArgHisIleHisIleValAspGlyA 198
|||||
101 GAACGTGCTGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 52

```

198 spArgLeuIleArgAspProIleProGluIleGlnIysValGluArgphe 214
|||||
51 ACCGGGTCATCGAGGAGAGGCTTCCCTAAAGATCCAAAAAGTCCAGGCTTC 2

```

seq_name: gb_est1:AA407647
seq_documentation_block:
LOCUS      AA407647          362 bp      mRNA      linear      EST 26-AUG-1998
DEFINITION musculus cDNA clone c0004A09 5', mRNA sequence
ACCESSION  AA407647
VERSION    AA407647.1  GI:2065848
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 362)
AUTHORS   Ko, M.S.H., Threat, T.A., Horton, J.H., Wang, X., Cui, Y., Wang, X., Pryor
          J.E., Paris, J.J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and
          Nakashima, H.
TITLE     Systematic analyses of mouse genes expressed in embryo implantation
          site
JOURNAL    Unpublished (1997)
COMMENT    Other_ESTs: EST00932
Contact: Ko MSH
Center for Molecular Medicine and Genetics
Wayne State University
5047 Gullen Mall, Detroit, MI 48202
Tel: 3135776708
Fax: 3135776200
Email: msk@cmb.biosci.wayne.edu
Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
            source
              1..362
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="ATCC (inhost):1363251"
                /db_xref="taxon:10090"
                /clone="C0004A09"
                /clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA
                library"
                /sex="unknown"
                /tissue_type="ectoplacental cone"
                /dev_stage="embryonic day 7.5 postconception"
                /lab_host="DH10B"
                /note="Organ: embryo; Vector: pSPBET1 (Life Technologies);
                Site 1: Salt; Site 2: Not 1; Total RNAs were extracted from
                ectoplacental cone of 7.5-dpc embryos. The
                double-stranded cDNA was synthesized from total RNAs with
                an Oligo(dT) primer. The library was constructed by Minoru
                S. H. Ko."
BASE COUNT 105 a 83 c 77 g 97 t
alignment_scores:
  Quality: 155.00      Length: 27
  Ratio: 5.741        Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-557-262-2_COPY_250_276 x AA407647
Align seq 1/1 to: AA407647 from: 1 to: 362
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyGly 17
12 TTCCTACTTTACAAACCAACGCGCTTCTATTCGCTTCGCGACAGCTGCGAA 61
17 sAspArgCysLeuHisGluSerLysGlyArg 27
62 GCACCGCTGCTTACACGAGTCCAAAGGCCGG 92
seq_name: gb_est2:BG090452
seq_documentation_block:
LOCUS      BG090452          425 bp      mRNA      linear      EST 26-JAN 2001
DEFINITION mac06h02.y1 Soares mouse 3NBMS Mus musculus cDNA clone
IMAGE:3999195.5' similar to TR:035310 035310 HEPARAN SULFATE
GIAUCOSAM:NYL 3-O SULFOTRANSFERASE PRECURSOR , mRNA sequence.
ACCESSION  BG090452
VERSION    BG090452.1  GI:12573015
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 425)
AUTHORS   NCI-CCRP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbts@mail.nih.gov
          This clone is available royalty-free through LLNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          MGI:11496923
Seq primer: -40RP from Gibco
High quality sequence stop: 378.
FEATURES   Location/Qualifiers
            source
              1..425
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:3999195"
                /clone_lib="Soares mouse 3NBMS"
                /sex="male"
                /tissue_type="Spleen"
                /dev_stage="4 weeks"
                /lab_host="DH10B"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker. Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
                was primed with a Not 1 - oligo(dT) primer [5',
                TGTTACCAATCTGAAGTGGAGCGCGCGCTGTTTTTTTTTTTTTTTTT
                3']; double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not 1 and cloned into the Not 1
                and Eco RI sites of the modified pT7T3 vector. RNA
                provided by Dr. Bertrand Jordan. Library went through
                three rounds of normalization, and was constructed by
                Bento Soares and M. Fatima Bonaldo."
BASE COUNT 104 a 124 c 93 g 104 t
alignment_scores:
  Quality: 155.00      Length: 27
  Ratio: 5.741        Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-557-262-2_COPY_250_276 x BG090452
Align seq 1/1 to: BG090452 from: 1 to: 425
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyGly 17
12 TTCCTACTTTACAAACCAACGCGCTTCTATTCGCTTCGCGACAGCTGCGAA 241
17 sAspArgCysLeuHisGluSerLysGlyArg 27
242 GCACCGCTGCTTACACGAGTCCAAAGGCCGG 272
seq_name: gb_est2:W62484
seq_documentation_block:
LOCUS      W62484          490 bp      mRNA      linear      EST 07-JUN 1996
DEFINITION md72c12.r1 Soares mouse embryo NME13.5 14.5 Mus musculus cDNA
clone IMAGE:373942 5' similar to PIR:A49733 A49733
glycosaminoglycan N-acetylglucosaminyl N-deacetylase ;, mRNA

```




```
sequence.
W62484
VERSION W62484.1 GI:1369243
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 490)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:235374
Seq primer: ETPrimer
High quality sequence stop: 344.
Location/Qualifiers
1. .490
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:373942"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue.type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pTT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pTT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 127 a 129 c 112 g 122 t
ORIGIN
alignment_scores:
Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-557-262-2_COPY_250_276 x W62484 ..
Align seg 1/1 to: W62484 from: 1 to: 490
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLy 17
|||||
196 TTCTACTTTAAACAAAACCAAGGCGCTTCTACTGCGTGGGACAGTGGCAA 245
|||||
17 SASPAQcysLeuHisGluSerLysGlyArg 27
|||||
246 GGACCGCTGCTTACAGGCTTCAAGGCGCG 276
|||||
seq_name. gb_est2:BM122797
seq_documentation_block:
LOCUS BM122797 492 bp. mRNA linear EST 01-FEB-2002
DEFINITION L0515C10-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA
Clone L0515C10 3', mRNA sequence.
ACCESSION BM122797
VERSION BM122797
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 492)
Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Pantano,S., Lim,M.K.
and Ko,M.S.H.
Systematic Analyses of NIA Mouse Newborn Heart cDNA Library
Unpublished (2001)
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6829, USA
Email: cdnaelgsun.grc.nia.nih.gov
Plate: L0515 row: C column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 492
POLYA=Yes.
Location/Qualifiers
1. 492
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:L0515C10-3"
/db_xref="taxon:10090"
/clone="L0515C10"
/clone_lib="NIA Mouse Newborn Heart cDNA Library"
/tissue.type="Newborn Heart"
/dev_stage="Newborn"
/lab_host="DH10B"
/notes="organ: heart; Vector: pSPORT1 (Invitrogen); Site_1:
SalI, Site_2: NotI; Mouse cDNA project by the Laboratory
of Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsn.grc.nia.nih.gov/GUNA).
Double-stranded cDNAs were synthesized with an oligo(dT)
primer [Invitrogen: 5'-
pACTAGTTCAGATGCGGCGCGCCCTTTTCTTTT-3'] from
24.9 microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to Lone-linker LL-Sal3 (Ref.
Development 127:1737-1749 (2000) [PMID: 10725249]),
purified by phenol/chloroform, and separated from free
linkers by Centricon 100. Then, the cDNAs were digested
with SalI and NotI enzymes, and cloned into SalI and NotI
site of pSPORT1 plasmid vector. The DH10B E. coli host was
transformed with ligation mixture by the chemical method.
The average insert size is about 1.8 kb. The library was
constructed by Yulan Piao (NIA)."
BASE COUNT 157 a 102 c 91 g 142 t
ORIGIN
alignment_scores:
Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-557-262-2_COPY_250_276 x BM122797/rev
Align seg 1/1 to reverse of: BM122797 from: 1 to: 492
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLy 17
|||||
488 TTCTACTTTAAACAAAACCAAGGCGCTTCTACTGCGTGGGACAGTGGCAA 439
|||||
```


OM of: US-09-557-262-2_COPY_250_276 to: N_Geneseq_032802 • Out_format pfs

Date: Jul 17, 2002 6:03 AM

About: Results were produced by the Genforce software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlip
-Q=/cqn2.1/hsptc1/spool/US09557262/anal_1607002_161205_25594/app_query.fasta.1.1568
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOF=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LCOEFL=0.000 -LCOEPEXT=0.000
-OGAP=0.4.500 -OGAPEXT=0.050 -XGAP=10.000 -XGAPEXT=0.500
-FGAP=6.000 -FGAPEXT=7.000 -YGAPOF=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATPIX=blotum62
-TRANS=human40.cdi -LIST=45 -DOCCALCN=200 -THR_SCORE=prt
-THR_MAX=100 -THR_MIN=0 -ALICN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09557262_@CNL_1_934 -NCTU=6 -ICPU=3 -LONGLOG
-DEVR_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THRRADS=1

Search information block:

Query: US-09-557-262-2_COPY_250_276

Query length: 27

Database: N_Geneseq_032802.*

Database sequences: 1736436

Database length: 85845721

Search time (sec): 742.980000

score_list.

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA37245 +	155.00	414.25	1.0e-14	16		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA65047 -	152.00	411.67	1.4e-14	9		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA67978 -	152.00	411.67	1.4e-14	9		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA35017 -	152.00	411.67	1.4e-14	9		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA16371 -	152.00	411.67	1.4e-14	9		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA42122 -	152.00	411.67	1.4e-14	9		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA14186 +	152.00	408.26	2.2e-14	13		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA37246 +	152.00	408.26	2.2e-14	13		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA46683 +	152.00	408.26	2.2e-14	13		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA26623 +	114.00	303.99	1.4e-08	8		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA34207 +	97.50	262.83	2.7e-06	5		
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA37250 +	97.50	262.83	2.7e-06	5		
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA37247 +	87.50	221.00	0.0006	19		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA52137 +	87.50	217.27	0.0009	20		
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA37249 +	85.50	214.92	0.0013	20		
/SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA37248 +	85.50	213.63	0.0015	23		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA18170 +	85.50	211.72	0.0019	2		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA52321 +	75.50	190.48	0.0284	11		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA47401 +	75.50	186.11	0.0518	21		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA113431 +	71.50	180.73	0.1018	1		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA113430 +	71.50	180.14	0.5113	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ53903 +	70.50	165.47	0.7206	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA54883 +	70.50	158.43	1.78	82		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA11309 +	59.50	134.15	40.01	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA12624 +	58.00	116.22	3.98	14		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA67711 +	57.50	149.83	5.46	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAQ5118 +	57.00	145.31	9.56	64		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA47920 +	56.00	131.20	58.36	2		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA60641 +	54.50	125.10	127.67	2		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA60641 +	54.00	139.69	19.65	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA77282 +	53.50	115.50	432.59	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA77283 +	53.50	115.50	432.59	4		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA56134 +	53.50	111.93	691.15	6		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA159415 +	52.00	124.95	130.14	1		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA161201 +	52.00	124.64	135.40	1		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA52142 +	52.00	122.98	167.68	1		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA68470 +	52.00	118.00	279.08	21		
/SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA65014 +	52.00	114.00	279.08	21		

alignment_scores.

Quality	Length
155.00	27
741	0
Percent Similarity: 100.000	Percent Identity: 100.000

/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA64160 + 52.00 115.00 272.08
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA60069 + 52.00 118.70 289.97
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA67924 + 52.00 118.13 312.19
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA64780 + 52.00 118.06 314.85
/SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA589455 + 52.00 117.06 357.98

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA37245

seq_documentation_block:

ID_AA37245_standard; PNA: 1685 BP.

XX AA37245;

XX 20-JUL-1999 (first entry)

XX Mouse 3-OST-1 encoding DNA.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase, anticoagulant;
XX saccharide, glycosaminoglycan, proteoglycan, antithrombin; 3-OST;
XX heparan sulfate, thrombotic disorder; deep vein thrombosis;
XX pulmonary embolism; coagulation enzyme inactivation; ss.

XX Mus musculus.

XX W99920085.A2

XX 06-MAY-1999.

XX 23-OCT-1998; 98W0-052597.

XX 31-OCT-1997; 97US-0065437.

XX 24-OCT-1997; 97US 0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Fritz GIMS, Liu J, Rosenberg PD, Schwartz EJ, Shwartz NM;

XX Zhang L;

XX WPI, 1999-312968/26.

XX P-ESDE; AAY17952.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related
XX polynucleotides

XX Claim 5: Page 1: 72, 95pp, English.

XX The invention relates to nucleic acid molecules (AA37245-X37250)

XX encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferase (3-OSTs).

XX The 3-OST proteins can be used for 3-O sulfating saccharide residues

XX within a preparation of glycosaminoglycan or proteoglycan

XX polysaccharides. 3-OST-1 can be used for enriching the

XX antithrombin binding fraction in a preparation of heparan sulfates (HS).

XX 3-OST 1 can also convert HS proteoglycan anticoagulant precursor to the

XX HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic

XX function) can be used to determine partial sequence information for

XX complex polysaccharides. The 3-OST proteins, genes and antibodies are

XX also especially used for diagnosis of disorders involving HS biosynthesis. 3-OSTs

XX are used to generate anticoagulant pentasaccharides, which may

XX be used to treat thrombotic disorders such as deep vein thrombosis and

XX pulmonary embolism. Coagulation enzyme inactivation by antithrombin is

XX enhanced by complexing of antithrombin with endothelial cell surface HS

XX proteoglycans. This is responsible for the non thrombotic properties

XX of blood vessels. The present sequence represents a mouse 3-OST-1

XX encoding DNA

XX Sequence 1685 bp, 422 A, 426 C, 382 G, 455 T, 0 other;


```

alignment_scores:
  CC at ftp.wipo.int/pub/published_pct_sequences.
  XX
  SQ Sequence 924 BP: 169 A; 252 C; 310 G; 193 T; 0 other;

alignment_scores:
  Quality: 152.00 Length: 27
  Ratio: 5.630 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
  US-09-557-262-2_COPY_250_276 x ABA35017/rev
  Align seq 1/1 to reverse of: ABA35017 from: 1 to: 924

  1 PheTyrPheAsnLysThriGlyCysLeuArgAspSerGlyLy 17
  170 TTCTACTTTTAAACAAACCAAGGGCTTTTACTGCTTGGGACACGGGCG 121

  17 saspArgCysLeuHisGluSerLysGlyArg 27
  120 GGACCGCTGCTTACATGAGTCAAGGCGG 90

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA35017
seq_documentation_block:
  ID ABA35017 standard; DNA: 924 BP.
  XX
  AC ABA35017;
  XX
  DT 23-JAN-2002 (first entry)
  DE
  DE Human; gene expression; heart; microarray; vascular system; probe;
  KW cardiovascular disease, hypertension, cardiac arrhythmia;
  KW congenital heart disease; ss.
  XX
  OS Homo sapiens.
  XX
  PN WO200157274-A2.
  XX
  PD 09-AUG-2001.
  XX
  PF 30-JAN-2001; 2001WO-US00666.
  XX
  PR 04-FEB-2000; 2000US-0180312.
  PP 26-MAY-2000; 2000US-0207456.
  PR 30-JUN-2000; 2000US-0608408.
  PR 03-AUG-2000; 2000US-0632366.
  PR 21-SEP-2000; 2000US-0234687.
  PR 27-SEP-2000; 2000US-0236359.
  PR 04-OCT-2000; 2000US-0024263.
  XX
  PA (MOLE-) MOLECULAR DYNAMICS INC.
  XX
  PI Penn SG, Hanzel DK, Chen W, Rank DR;
  XX
  DR WPI; 2001-488899/53.
  XX
  PT Single exon nucleic acid probes for analyzing gene expression in human
  PT hearts -
  XX
  PS Claim 4; SEQ ID No 13483; 530pp; English.
  XX
  CC The present invention relates to single exon nucleic acid probes for
  CC measuring human gene expression in a sample derived from human heart. The
  CC present sequence is one such probe. The probes may be used for
  CC predicting, measuring and displaying gene expression in samples derived
  CC from the human heart via microarrays. By measuring gene expression, the
  CC probes are useful for predicting, diagnosing, grading, staging,
  CC monitoring and prognosing diseases of the human heart and vascular system
  CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
  CC congenital heart disease.
  CC Note: The sequence data for this patent did not form part of the printed
  CC specification, but was obtained in electronic format directly from WIPO

```

Example 4: SEQ ID NO: 16462; 650pp + Sequence Listing; English.
 The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 brain. They can be used to measure gene expression in brain cell samples,
 which may enable the diagnosis and improved treatment of nervous system
 diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 epilepsy and cancers. The present sequence is one of the probes of the
 invention.

CC expression of nucleic acid molecules in response to a stimulus, e.g. a
 CC chemical, drug or growth factor (especially epidermal growth factor),
 CC radiation, stress or a pathogen. The methods can also be used to
 CC determine co-regulated genes that can be potential targets for drug
 CC discovery.

XX
 SQ Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:

Quality: 152.00 Length: 27
 Ratio: 5.630 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-2_COPY_250_276 x AA236687

Align seq 1/1 to: AA236687 from: 1 to: 1305

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyIy 17
 |||||
 854 TTCTACTTTAA:AAAACCAAGGGTTTAA:TTGTTTGGGA:ATPAGGGG 903

17 SASpArgCysLeuHisGluSerLysGlyArg 27

|||||

904 GCACCGCTGCTTACATGAGTCCAAAGCCCG 934

seq_name: /SID55/grgdata/4-us-seq/geneseq-nmb1/NA2001R.1AT:ABL26623

seq_documentation_block:

ID: ABL26623 standard; DNA; 897 BP.

XX

XX ABL26623;

XX

XX 26-MAR-2002 (first entry)

XX

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31342.

XX Drosophila; developmental biology; cell signalling, insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX

XX WO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -

XX

XX Claim 1: SEQ ID NO 31342; 21pp - Sequence Listing, English.

XX

XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX sequences (ABH57737-ABH72072).

XX The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 897 BP; 183 A; 233 C; 295 G; 186 T; 0 other;

alignment_scores:

Quality: 114.00 Length: 27
 Ratio: 4.560 Gaps: 0
 Percent Similarity: 92.593 Percent Identity: 74.074

alignment_block:

US-09-557-262-2_COPY_250_276 x ABL26623

Align seq 1/1 to: ABL26623 from: 1 to: 897

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyIy 17
 |||||
 703 TTTTACTTCAATGAGACCAAGCGCTTCTACTGTCGCTATGACACCG 752

17 SASpArgCysLeuHisGluSerLysGlyArg 27

|||||

753 GATGCTGATTTGGAGAGAGAAAGGGAGG 783

seq_name: /SID55/grgdata/4-us-seq/geneseq-nmb1/NA2001R.1AT:ABL26623

seq_documentation_block:

ID: ABL26623 standard; DNA; 348; BP.

XX

XX ABL26623;

XX

XX 26-MAR-2002 (first entry)

XX

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31342.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX

XX WO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -

XX

XX Claim 1: SEQ ID NO 31339; 21pp - Sequence Listing, English.

XX

XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX sequences (ABH57737-ABH72072).

XX The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX

CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a human 3-OST-4
 CC encoding DNA.

XX
 SQ Sequence 3658 BP; 771 A, 1085 C, 1167 G; 695 I, 0 other;

alignment_scores:
 Quality: 97.50 Length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 66.667

alignment_block:
 US-09-557-262-2_COPY_250_276 x AAX37250 ..

Align seg 1/1 to: AAX37250 from: 1 to: 3658

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArg.....As 14
 |||||
 2002 TCTATTTCGAAAGAAAGAAAGAGTTTCTGCTTAAAGAAAGAAAGAAAGA 2051
 |||||
 14 pSerGlyLysAspArgCysLeuHisGluSerLysGlyArg 27
 |||||
 2052 CAGCAGTCGCGCGAGAGTCTGTTAGGAAAGAAAGAAAGAGTGG 2091

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: AAX37247

seq_documentation_block:

ID AAX37247 standard; DNA, 1951 BP.

XX AAX37247,

DE 20-JUL-1999 (first entry)

DE Human 3-OST-2 encoding DNA.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 KW heparan sulfate; thrombotic disorder; deep vein thrombosis;
 KW pulmonary embolism; coagulation enzyme inactivation; ss.

OS Homo sapiens.

XX W09922005-A2.

XX 06-MAY-1999.

PD 23-OCT-1998; W0991-US22597.

XX 31-OCT-1997; 97US-0065437.

PR 24-OCT-1997; 97US-0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Fritze JMS, Liu J, Fosenberg R, Schwartz JJ, Shwartz NW,
 PI Zhang L;

DR WPL; 1999-312968/26.

DR P-PSDB; AAY17064.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related
 PT polynucleotides

XX Claim 5; Page 76-78; 95pp; English.

XX The invention relates to nucleic acid molecules (AAX37245-X7250)

CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs)

CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues
 CC within a preparation of glycosaminoglycan or proteoglycan
 CC polysaccharides. 3-OST-1 can be used for enriching the
 CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).
 CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 CC HS anticoagulant (disac). The 3-OSTs (optionally lacking enzymatic
 CC function) can be used to determine partial sequence information for
 CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a human 3-OST-2
 CC encoding DNA.

XX Sequence 1951 BP; 424 A; 603 C; 490 G; 434 T; 0 other;

alignment_scores:
 Quality: 87.50 Length: 30
 Ratio: 3.804 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 60.000

alignment_block:
 US-09-557-262-2_COPY_250_276 x AAX37247 ..

Align seg 1/1 to: AAX37247 from: 1 to: 1951

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLy 17
 |||||
 979 TCTATTTCGAAAGAAAGAGTTTCTGCTTAAAGAAAGAAAGAAAGA 1028
 |||||
 17 Asp.....ArgCysLeuHisGluSerLysGlyArg 27
 |||||
 1029 GAGCAGTCGCGCGAGAGTCTGTTAGGAAAGAAAGAAAGAGTGG 1068

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2801A.DAT: AAS21347

seq_documentation_block:

ID AAS21347 standard; cDNA; 2845 BP.

XX AAS21347;

XX 24-OCT-2001 (first entry)

DE Human cDNA sequence encoding for pK05004 polypeptide.

XX Human secretory and transmembrane, PRO, mammalian, cancer, lung,
 KW breast, prostate, cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage, ear, proliferation, glucose, free fatty acid, skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

OS Homo sapiens.

XX WC200140466-A2.

XX 07-JUN-2001.

PD 01-DEC-2000; 2000W0-US342678.

XX 01-DEC-1999; 99W0-US28301.

PR 01-DEC-1999; 99W0-US28634.

PR 02-DEC-1999; 99W0-US28551.

PR 02-DEC-1999; 99W0-US28564.

PR 02-DEC-1999; 99W0-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99W0-US30095.

PR 20-DEC-1999; 99W0-US30911.

PR 30-DEC-1999; 99W0-US30999.

PR 06-JAN-2000; 2000W0-US00277.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that proper record-keeping is essential for transparency and accountability, particularly in financial matters. The text outlines various methods for organizing and storing data, including digital databases and physical filing systems. It also mentions the need for regular audits and reviews to ensure the integrity of the information.

2. The second section focuses on the role of communication in achieving organizational goals. It highlights that effective communication is a key factor in building a cohesive team and fostering a positive work environment. The text provides practical advice on how to improve communication skills, such as active listening, clear articulation of ideas, and the use of appropriate communication channels. It also discusses the importance of maintaining open lines of communication between all levels of the organization.

3. The third part of the document addresses the challenges of managing time and resources efficiently. It notes that time is a valuable asset, and its effective management is crucial for the success of any project or organization. The text offers strategies for prioritizing tasks, delegating responsibilities, and avoiding procrastination. It also touches upon the importance of resource allocation and the need to make the most of the available budget and personnel.

4. The final section discusses the importance of continuous learning and professional development. It states that in a rapidly changing world, individuals and organizations must stay updated with the latest trends and technologies. The text encourages the pursuit of new knowledge and skills through various means, such as attending workshops, conferences, and taking courses. It also emphasizes the value of mentorship and peer learning in accelerating professional growth.


```

DEFINITION      W706c11.x1 NCL_CGAP_G06 Homo sapiens cDNA clone IMAGE:2480756 3'
                  similar to TR014792 014792 HEPARAN SULFATE 3-O-SULFOTRANSFERASE-1
PRECUSOR.       17 mRNA sequence.
ACCESSION       AI971626
VERSION         AI971626.1 GI:5768452
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE       1 (bases 1 to 651)
AUTHORS        NCL_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL        Tumor Gene Index
COMMENT        Unpublished (1997)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapbs@mail.nih.gov
                Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
                R. Emmert-Buck, M.D., Ph.D.
                cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
                Bonaldo, Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCL_CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www.bio.llnl.gov/bbrp/image/image.html
                Insert Length: 1020 Std Error: 0.00
                Seq primer: -400P from Gibco
                High quality sequence stop: 456.
FEATURES        location/Qualifiers
                 1..651
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone IMAGE:2480756"
                 /clone_lib="NCL_CGAP_G06"
                 /tissue_type="pooled germ cell tumors"
                 /lab_host="pH10H"
                 /note="Vector: pTZ19-pac (Pharmacia) with a modified
                 polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA
                 from the normalized library NCL_CGAP_G04 was prepared, and
                 ss circles were made in vitro. Following HAP purification,
                 this DNA was used as tracer in a subtractive hybridization
                 reaction. The driver was PCR-amplified cDNAs from a pool
                 of 5,000 clones made from the same library (clones
                 1257096-1258631, 1459054-1470983, and 1475592-1476743).
                 Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      151 a 211 c 158 g 141
ORIGIN
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyAr 17
|||||
466 TTATATTATTAACAAAACAAAGGCTTTTATGCTCTGGGGAAAGGGGG 515
|||||
17 qAspArgCysLeuHisGlnSerLysGlyArg 27
|||||
516 GCACCGCGCTTACATACAGTCCAAAGCGCGCT 546
|||||
seq_name: qb_est2:BG820537
seq_documentation_block:
LOCUS          BG820537
DEFINITION    602782784F1 NCL_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933675
5', mRNA sequence.
ACCESSION     BG820479
VERSION       BG820479
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE     1 (bases 1 to 788)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
                Email: cgapbs@mail.nih.gov
                Tissue Procurement: David N. Louis, M.D.
                cDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM10861 row: q column: 20
                High quality sequence stop: 783.
FEATURES      location/Qualifiers
               1..783
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone IMAGE:4933675"
               /clone_lib="NCL_CGAP_Brn67"
               /tissue_type="anaplastic oligodendroglioma with ip/19q
               loss"
               /lab_host="pH10H (11 phage-resistant)"
               /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: Not 1;
               Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
               Average insert size 2.3 kb. Constructed by Life
               Technologies. Note: this is a NCL_CGAP library."
BASE COUNT    203 a 227 c 175 g 178 t
ORIGIN
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyAr 17
|||||
435 TTCTAGTTTAAACAAAACAAAGGCTTTTATGCTCTGGGGAAAGGGGG 484
|||||
17 qAspArgCysLeuHisGlnSerLysGlyArg 27
|||||
487 GGAAGCTGTGTTATATGAGTCCAAAGCGCGG 515
|||||
seq_name: qb_est2:BG820479
seq_documentation_block:
LOCUS          BG820479
DEFINITION    602782784F1 NCL_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933676
5', mRNA sequence.
ACCESSION     BG820479
VERSION       BG820479
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE     1 (bases 1 to 788)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.

```


TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1AM10861 row: q column: 21
High quality sequence stop: 777.

FEATURES source

1..788
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:493676"
/clone_lib="MCL CGAP_Bn67"
/tissue_type="anaplastic oligodendrogloma with lp/19q
loss"
/lab_host="DH10B (M1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-Script6; Site: 1, Not 1;
Site 2, SalI. Cloned unidirectionally. Primer: oligo d;
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: This is a MCL CGAP Library."
BASE COUNT 209 a 224 c 174 g 181 t
ORIGIN

alignment_scores:

Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-557-262-4_copy_246_272 x B3820479 ..

Align seq 1/1 to: B3820479 from: 1 to: 788

1 PhcTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
435 TTCTACTTTAAVAAAACCAAGGGCTTTACTGCTGGTGGTAAATGGGG 484

17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||
485 GGACCGCTGCTTACATGAGTCCCAAGGGCGG 515

seq_name: qb_est1:AI154059

documentation_block:

US AI154059 283 bp mRNA linear EST 30-Sep-1998
DEFINITION ud58f01.r1 Soares_NMPu Mus musculus cDNA clone IMAGE:1450105 5'
similar to TP 035310 035310 HEPAPAN SULFATE PRECURSOR 3'-UTR-SAMINYL
3'-O-SULFOTRANSFERASE-1 PRECURSOR .: mRNA sequence.

AI154059

ACCESSION AI154059.1 GI:3682528

VERSION EST.

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata, Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 283)
REFERENCE Marra M., Hillier L., Allen M., Rowles M., Bieriich N., Bubaque T.,
Geisler S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
Schellenberg K., Slepoe M., Tan F., Underwood K., Moore R.,
Theising B., Wylie T., Lennon G., Soares B., Willson P. and
Waterston R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
Washington University School of Medicine

4444 Forest Park Parkway, Box #501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:923421

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 1.

FEATURES

source

1..283
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1450105"
/clone_lib="Soares_NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus, and was then primed with a Not I -
oligo(dT) primer. Double stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Ronaldo."
BASE COUNT 80 a 30 c 64 g 59 t
ORIGIN

alignment_scores:

Quality: 152.90 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 95.296

alignment_block:

us-09-557-262-4_copy_246_272 x AI154059 ..

Align seq 1/1 to: AI154059 from: 1 to: 283

1 PhcTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
98 TTCTACTTTAAACAAACCAAGGGCTTCTACTGCCGCGCACAGTGGCAA 147

17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||

148 GGAAGGCTGCTTACATGAGTCCCAAGGGCGG 178

seq_name: qb_est2:BB850108

documentation_block:

US BB850108 291 bp mRNA linear EST 26-Sep-2000
DEFINITION ud58f01.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:1416472
5' similar to TP 035310 035310 HEPAPAN SULFATE PRECURSOR .: mRNA sequence.

3 O-SULFOTRANSFERASE PRECURSOR .: mRNA sequence.

ACCESSION BB850108

VERSION BB850108.1 GI:10408447

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata, Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 291)

REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```

source
1. .359
/oranism-"Mus musculus"
/strain-"C57H1/6J"
/db_xref-"taxon:10090"
/clone-"IMAGE:1999170"
/musculi_lib-"Soares mouse 3NMS"
/sex-"male"
/tissue_type-"Spleen"
/dev_stage-"4 weeks"
/lab_host-"DH10B"
/note-"Vector: pT7-3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(3T) primer [5]. TGTACCAATGAGTAGTGAGCGGCGGGTGTGTTTTTTTTTTTTTTT [5]; double-stranded cDNA was ligated to Eco RI adaptors (pharmacica), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7-3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Falima Bonaldo."
BASE COUNT      93 a   107 c    77 g    82 t
ORIGIN

alignment_scores:
Quality:     152.00          Length:       27
Ratio:       5.630           Gaps:         0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-ES1-262-4_gary_246_272 x B090529 ..
Align seq 1/1 to: H0090529 from: 1 to: 359

1 PctTyrPheAsnArgHisPheGlySerPheTyCysAlaArgAspSerGlyAr 17
|||||
192 TTCTACTTTAA'AAAAT'AAAGAGATTACTGGTTCCTGA'ACTGTA 241

17 gspATgCgcScchHicGluScrTySGLYArq 27
|||||
242 GCACGCCTGCTTACACACACTCAAGGCCGCC 272

seq_name: db_est1:AA407647

seq_documentation_block:
LOCUS      AA407647              362 bp mRNA linear EST 26 Aug 1998
DEFINITION ESF06933 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus
musculus cDNA clone C0004A09 5' mRNA sequence.
ACCESSION  AA407647
VERSION    AA407647.1 GI:2065848
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurodonthii; Muridae; Murinae; Mus
1 (bases 1 to 362)
K.O.M.S.H., Throat,T.A., Horton,J.H., Wang,X., Gul,Y., Wang,X., Pryor
,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and
Nakashima,H.
Systematic analyses of mouse genes expressed in embryo implantation
site
JOURNAL    Unpublished (1997)
COMMENT    Other_FSTS: ESF00932
Contact: KO MSH
Center for Molecular Medicine and Genetics
Wayne State University
5047 Gullen Mall, Detroit, MI 48202
Tel: 3135776708
Fax: 3135776200
Email: mskocmb.biosci.wayne.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..362
FEATURES
```

```

MG1:1092284
Seq primer: -40RP from Gibco
High quality sequence stop: 269.
FEATURES
SOURCE
    location/Qualifiers
        1..291
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:1416472"
            /clone_lib="Soares mouse 3NDMS"
            /sex="male"
            /tissue_type="Spleen"
            /dev_stage="4 weeks"
            /lab_host="DB10H"
            /note="Vector: p7130-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTAATACCTCAATTGAGAGGGGCGTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p713 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
BASE COUNT      68 a   92 c   69 g   62 t
ORIGIN
alignment_scores:
    Quality: 152.00
    Ratio: 5.630
    Percent Similarity: 100.000
    Percent Identity: 96.296
alignment_block:
DS-09-557-262_4_C-IV_246_272 x H0950128
Align seq 1/1 to: BE85010R from: 1 to: 291
1 PhcYTPPhcAsnLYSThrLYSclyPheCYrCYStLeuAArgSPScGLyAr 17
|||||
192 TTTTAACTTTAAACAAAACAACCCCTCTTACTGTGTCGCCCATACAGCCCAA 241

17 qAspArqCysLeuHisGluSerLysGlyArg 27
|||||
242 GCACGTCTGCTTACACGCAGACGACAAAGGCCCG 272

seq_name: qE_cst2.HG090529
seq_documentation_block:
LOCUS       BG090529               350 bp          mRNA           linear     EST 26-JAN-2001
DEFINITION  mac07402.y1 Soares mouse 3NDMS Mus musculus cDNA clone
IMAGE: 3999170.57 similar to TR_035310 G35310 HEFARN SU/FATE
GLUCOSAMINYL 3-O-SULFOTRANSFERASE PRECURSOR ;, mRNA sequence.
ACCESSION   BG090529
VERSION     BG090529.1 GI:12573092
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Solunguathi; Muridae; Mus;
1 (bases 1 to 359)
NOT REAP http://www.ncbi.nlm.nih.gov/ncbiqap
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Surasterq, Ph.D.
Email: eqabbs@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:1496898
Seq primer: -40RP from Gibco
High quality sequence stop: 347.
location/Qualifiers

```



```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:373942"
/clone_lib="Soares mouse embryo NME13.5 14.5"
/issue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/Note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker, Site1, Not I, Site2, Eco RI, 1st strand cDNA
was primed with a Not I, cigo(3T) primer 15'
TGTTCACCAATGAAAGAGGAGCGGCGGAAATTTTTTTTTTTTTTTT
T 3', on equal amounts of mRNA provided by Minora KO, Wayne
State Univ., from 2 l; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT      127 a  129 c  112 g  122 t
ORIGIN

```

```

alignment_scores:
  Quality: 152.00      Length: 27
  Ratio: 5.630        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

```

alignment_block:

US-09-557-262-4_copy_246_272 x W62484

Align seq 1/1 to: W62484 from: 1 to: 490

```

1 PhetYrheASnlySbhrYsGlypHeYrCysLeuArgASpScrGlyAr 17
|||||
196 TTCTACITTAACAAAGCAAGAGTGTATATGCTGCGGAGATAGTGGCAA 245

```

```

17 qASpArGcYsLeuHisGluSerYsGlyArq 27
|||||
246 GCACCGCTGCTTACACGAGTCCAAAGCGCGG 276

```

seq_name: qb_est2:BM122797

seq_documentation_block:

```

LOCUS      BM122797              492 bp      mRNA      linear      EST 01-FEB-2002
DEFINITION clone L0515C10-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA
ACCESSION  BM122797.1  GI:17104565
VERSION     BM122797
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

```

REFERENCE 1 (bases 1 to 492)

Plato,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Pantano,S., Lim,M.K. and Ko,M.S.H.

Systematic Analyses of NIA Mouse Newborn Heart cDNA Library

Unpublished (2001)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@qsun.grc.nia.nih.gov

Plate: L0515 row: C column: 10

Seq primer: -21M13 Forward

High quality sequence stop: 492

POLYA-Yes.

Location/Qualifiers

1..492

Organism="Mus musculus"

/strain="C57BL/6J"

FEATURES

Source

```

/db_xref="niaEST:L0515C10-3"
/db_xref="taxon:10090"
/clone="L0515C10"
/clone_lib="NIA Mouse Newborn Heart cDNA Library"
/tissue_type="Newborn Heart"
/dev_stage="Newborn"
/lab_host="DH10B"
/Note="Organ: heart; Vector: pSPORT1 (Invitrogen); Site 1:
Site2: NotI; Mouse cDNA project by the Laboratory
of Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://lqsun.grc.nia.nih.gov/cDNA).
Double-stranded cDNAs were synthesized with an oligo(dT)
primer (Invitrogen): 5'-
pACTAGTCTAGATGCGGAGCGGCGGAAATTTTTTTTTTTT 3' from
24.9 microgram of total RNA, treated with 14 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to Bam linker (B-Sal3 (Ref.
Development 127:1737-1749 (2000) PMID: 10725249)),
purified by phenol/chloroform, and separated from free
linkers by Centricon 100. Then, the cDNAs were digested
with SalI and NotI enzymes, and cloned into SalI and NotI
site of pSPORT1 plasmid vector. The DH10B E. coli host was
transformed with ligation mixture by the chemical method.
The average insert size is about 1.8 kb. The library was
constructed by Yulan Plato (NIA)."
BASE COUNT      157 a  102 c  91 g  142 t
ORIGIN

```

alignment_scores:

```

  Quality: 152.00      Length: 27
  Ratio: 5.630        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

```

alignment_block:

US-09-557-262-4_copy_246_272 x BM122797/rev

Align seq 1/1 to reverse of: BM122797 from: 1 to: 492

```

1 PhetYrheASnlySbhrYsGlypHeYrCysLeuArgASpScrGlyAr 17
|||||
488 TTCTACITTAACAAAGCAAGAGTGTATATGCTGCGGAGATAGTGGCAA 490

```

```

17 qASpArGcYsLeuHisGluSerYsGlyArq 27
|||||
438 GCACCGCTGCTTACACGAGTCCAAAGCGCGG 408

```

seq_name: qb_est2:BM124136

seq_documentation_block:

```

LOCUS      BM124136              504 bp      mRNA      linear      EST 01-FEB-2002
DEFINITION clone L0535G06-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA
ACCESSION  BM124136
VERSION     BM124136.1  GI:17107904
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

```

REFERENCE 1 (bases 1 to 504)

Plato,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Pantano,S., Lim,M.K. and Ko,M.S.H.

Systematic Analyses of NIA Mouse Newborn Heart cDNA Library

Unpublished (2001)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@qsun.grc.nia.nih.gov

Plate: L0535 row: G column: 06

Seq primer: -21M13 Forward

High quality sequence stop: 504

FEATURES
source
 PolYA-Yes.
 Location/Qualifiers
 1..504
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="EnaEST:L053506.3"
 /db_xref="taxon:10090"
 /clone="L053506"
 /clone_lib="NIA Mouse Newborn Heart cDNA library"
 /tissue_type="Newborn Heart"
 /dev_stage="Newborn"
 /lab_host="DH10B"
 /note="Organ: heart; Vector: pSPOR1 (Invitrogen); Site: 1; Salt: 2; NotI Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://www.genet.nia.nih.gov/inaa) Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-
 pGACTAGTCTATGATGCAAGCGGCTTCTTTTCTTTT-3'] from 24.9 microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol precipitation. The cDNAs were ligated to lone-linker BL Sal3 (Ref. Development 127:1737-1745 (2000) [PMID: 10735491]), purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were digested with SalI and NotI enzymes, and cloned into SalI and NotI site of pSPOR1 plasmid vector. The DH10B E. coli host was transformed with ligation mixture by the chemical method. The average insert size is about 1.8 kb. The library was constructed by Yulan Piao (NIA)."
BASE COUNT
 159 a 104 c 96 g 145 t
ORIGIN
 alignment_scores
 Quality: 152.00 Length: 27
 Ratio: 5.630 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 96.206
 alignment_block:
 US-09-557-262-4_Cory_246_272 x HM124136/rev ..
 Align seq 1/1 to reverse of: BM124136 from: 1 to: 504
 1 PheTyrPheAsnIlyThrIysGlyPheTyrCysLeuArgAspSerGlyArg 17
 488 TTCTACTTTTAAACAAACCAAGGCGCTCTACCTGCTGCGGACAGTGCAC 439
 17 qAspArgCysLeuHisSerIysGlyArg 27
 438 CGACCGCTGCTTACACGAGTCCAAAGCGCG 408
 seq_name: gb_est1:AA0407566
 seq_documentation_block:
 LOCUS AA0407566 504 bp mRNA linear EST 03-SEP-1996
 DEFINITION clone IMAGE:475268 5' similar to PIR:AA9733 AA9733
 glycosaminoglycan N acetylglucosaminyl H deacetylase ; mRNA sequence.
 accession AA041885
 version AA041885.1 GI:1520040
 keywords EST.
 source house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 507)
 REFERENCE
 AUTHORS
 Ko,M.S.H., Threlk,A., Horton,J.H., Wang,X., Gui,Y., Wang,X., Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and Nakashima,H.
 TITLE
 Systematic analyses of mouse genes expressed in embryo implantation site
 JOURNAL
 Unpublished (1997)
 COMMENT
 Other ESTs: EST00440

WashU-HIMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info:image.llnl.gov) for further information.

MCL:285012

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 490.

FEATURES

source

Location/Qualifiers
1..508
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:475268"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/issue_type="embryo"
/dev_stage="14.5,14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pUT73b-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
was primed with a Not 1 - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGGGGCGGCTGTGTGTGTGTGTGTGTGTGT
T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minora Ko, Wayne
State Univ., from 2 l, double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not 1 and
cloned into the Not 1 and Eco RI sites of the modified
pUT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Ronaldo."

BASE COUNT 136 a 134 c 114 g 124 t

ORIGIN

alignment_scores:

Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_COPY_246_272 x AAC41985
align seq 1/1 to: AAC41885 from: 1 to: 568
1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuArgAspSerGlyArg 17
|||||
196 TTC TAC TTA CAA A A A A C C A A G G G C T T C T A C T G C T C G G G A C A G T C G C A A 245

17 qAspArgCysLeuHisGlySerLysGlyArg 27

|||||

246 GGA CCG TGA TTA CAG GAG TCA AAG GAG G G G 276

seq_name: qb_est2:HE851982

seq_documentation_block:

LOCUS BE851982 555 bp mRNA linear EST 26-SEP-2000
DEFINITION w12e07.x1 Soares mouse mBMS Mus musculus cDNA clone IMAGE:3416484
3' similar to TR:035310.035310 HEPARAN SULFATE GLUCOSAMINYL
3'-O-SULFOTRANSFERASE PRECURSOR ?, mRNA sequence.

ACCESSION BE851982

VERSION BE851982.1 GI:10310321

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 555)

AUTHORS NC1-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

COMMENT

Tumor Gene Index
Unpublished (1997)
Other_FSTS: w12e07.y1
Contact: Robert Strausberg, Ph.D.
Email: rcgaps@mail.nih.gov
This clone is available royalty free through LLNL; contact the
IMAGE Consortium (info:image.llnl.gov) for further information.
MGI:1092296
High quality sequence stop: 474.

FEATURES

source

Location/Qualifiers
1..555
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3416484"
/clone_lib="Soares mouse mBMS"
/sex="male"
/issue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pUT73b-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
was primed with a Not 1 - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGGGGCGGCTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not 1 and cloned into the Not 1
and Eco RI sites of the modified pUT73 vector. RNA
provided by Dr. Norihiro Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Ronaldo."

BASE COUNT

173 a 111 c 114 g 157 t

alignment_scores:

Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_COPY_246_272 x BE851982/rev

Align seq 1/1 to reverse of: BE851982 from: 1 to: 555

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuArgAspSerGlyArg 17
|||||

486 TTC TAC TTA CAA A A A A C C A A G G G C T T C T A C T G C T C G G G A C A G T C G C A A 437

17 qAspArgCysLeuHisGlySerLysGlyArg 27

|||||

436 GGA CCG TGA TTA CAG GAG TCA AAG GAG G G G 406




```

sequence.
ACCESSION W62484
VERSION W62484.1 GI:1369243
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Pukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 490)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, J., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
TITLE Unpublished (1996)
JOURNAL
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty free through HMI; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:235374
Seq primer: EST primer
High quality sequence stop: 344.
FEATURES
source
1..490
Location/Qualifiers
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone_image="373942"
/clone_lib="Soares mouse embryo NME13 5 14 5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT73-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Eco RI; 1st strand cDNA
was primed with a Not 1 - oligo(dT) primer 15',
TGTTCACCAATCGAAGTGGAGCGCCGCCGAAATTTTTTTTTTTTTTTTTT
1 3', on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Mifera KO, Wayne
State Univ, from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not 1 and
cloned into the Not 1 and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 127 a 129 c 112 g 122 t
ORIGIN

alignment_scores:
Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-2_COPY_250_276 x W62484
Align seq 1/1 to: W62484 from: 1 to: 490
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyIly 17
|||||
196 TTCTACTTTTAAACAAACCAAGGGCTTCTACTGCGTGGCGACAGTGGCAA 245

17 SASPArcCysLeuIleGluSerLysGlyArc 27
|||||
246 GGACCGCTGCTTACACAGATCCAAAGCGCG 276

```

```

seq_name: qb_est2:BM122797
seq_documentation_block:
LOCUS BM122797 492 bp mRNA linear EST 01 FEB 2002
DEFINITION L0515C10-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA
clone L0515C10 3', mRNA sequence.
ACCESSION BM122797
VERSION BM122797.1 GI:17106565
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Pukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 492)
Piao, Y., Korqul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
and Ko, M.S.H.
TITLE Systematic Analyses of NIA Mouse Newborn Heart cDNA Library
JOURNAL Unpublished (2001)
COMMENT Contact: Howard R. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@nidsun.grc.nia.nih.gov
Plate: 0515 10w, C column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 492
POLYA-Yes.
FEATURES
Location/Qualifiers
1..492
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="niaEST:L0515C10-3"
/db_xref="taxon:10090"
/clone_image="L0515C10"
/clone_lib="NIA Mouse Newborn Heart cDNA Library"
/tissue_type="Newborn Heart"
/dev_stage="DH10B"
/lab_host="DH10B"
/notes="Organ: heart; Vector: pSPORT1 (Invitrogen); Site 1:
SalI; Site 2: NotI; Mouse cDNA project by the Laboratory
of Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igson.grc.nia.nih.gov/CDNA);
double-stranded cDNAs were synthesized with an oligo(dT)
primer [Invitrogen: 5'-
pCACTAGTCTAGATGCGAGCGCGGCTTTTTTTTTTTT-3'] from
24.9 microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to lunc-linker LL-Sal3 (ref.
development 127:1737-1749 (2000) [PMID: 10725249]),
purified by phenol/chloroform, and separated from free
linkers by centrifugation 100. Then, the cDNAs were digested
with SalI and NotI enzymes, and cloned into SalI and NotI
site of pSPORT1 plasmid vector. The DH10B E. coli host was
transformed with ligation mixture by the chemical method.
The average insert size is about 1.8 kb. The library was
constructed by Yulan Piao (NIA)."
BASE COUNT 157 a 102 c 91 g 142 t
ORIGIN

alignment_scores:
Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-2_COPY_250_276 x BM122797 rev
Align seq 1/1 to reverse of: BM122797 from: 1 to: 492
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyIly 17
|||||
488 TTCTACTTTTAAACAAACCAAGGGCTTCTACTGCGTGGCGACAGTGGCAA 439

```

17 sAspArqCysLeuHisGluSerLysGlyArg 27
 |||||
 438 GGACCGCTGCTTACACGAGTCCAAAGCGCG 408

seq_name: gb_est2:BM124136

seq_documentation_block:

LOCUS BM124136 504 bp mRNA linear EST 01-FEB-2002
 DEFINITION L0535G06-3 NIA Mouse Newborn Heart cDNA library Mus musculus cDNA
 clone L0535G06 3', mRNA sequence.

ACCESSION BM124136

VERSION BM124136 1 GI:17107904

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 504)

Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.

and K.O.M.S.H.

Systematic Analyses of NIA Mouse Newborn Heart cDNA Library

UNPUBLISHED (2001)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224 6830, USA

Email: cdnaelgusun.grc.nia.nih.gov

Plate: L0535 row: C column: 06

Seq primer: -21M13 Forward

High quality sequence stop: 504

POLYA-yes.

FEATURES

source

location/Qualifiers

1..504

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="niaEST:L0535G06-3"

/db_xref="taxon:10090"

/clone="L0535G06"

/clone_lib="NIA Mouse Newborn Heart cDNA Library"

/tissue_type="Newborn Heart"

/dev_stage="Newborn"

/lab_host="DH10B"

/note="Organ heart. Vector pSPORT1 (Invitrogen). Site 1-

SalI; Site 2: NotI; Mouse cDNA project by the laboratory

of Genetics, National Institute on Aging (NIA), Intramural

Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>).

Double-stranded cDNAs were synthesized with an Oligo(dT)

primer (Invitrogen): 5'-

PGACTAGTCTAGATCGAGCGCGGCGCTTTTCTTTT-3'}

from

24.9 microgram of total RNA, treated with T4 DNA

polymerase, and purified by ethanol-precipitation. The

cDNAs were ligated to lona-linker Li-Sal3 (Ref.

Development 127-1747-1749 (2000) [PMID: 10725240],

purified by phenol/chloroform, and separated from free

linkers by phenol extraction. Then, the cDNAs were digested

with SalI and NotI enzymes, and cloned into SalI and NotI

site of pSPORT1 plasmid vector. The DH10B E. coli host was

transformed with ligation mixture by the chemical method.

The average insert size is about 1.8 kb. The library was

constructed by Yulan Piao (NIA)."

BASE COUNT 159 a 104 c 96 g 145 t

ORIGIN

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Percent identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x BM124136/rev

Align seq 1/1 to reverse of: BM124136 from: 1 to: 504

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuAqAspSerGly 17
 |||||
 488 TTCTACTTTTAAACAAACCAAGCGCTTCTACAGCTCCAAAGCGCG 439

17 sAspArqCysLeuHisGluSerLysGlyArg 27

|||||

438 GGACCGCTGCTTACACGAGTCCAAAGCGCG 408

seq_name: gb_est1:AA407566

seq_documentation_block:

LOCUS AA407566

DEFINITION F5700931 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus
 musculus cDNA clone C0004A08 5', mRNA sequence.

ACCESSION AA407566

VERSION AA407566 1 GI:2065846

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 507)

Ko, M.S.H., Threalt, T.A., Horton, T.H., Wang, X., Wang, X., Pryor

, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and

Nakashima, H.

Systematic analyses of mouse genes expressed in embryo implantation

site

UNPUBLISHED (1997)

Contact: Ko MSH

Center for Molecular Medicine and Genetics

Wayne State University

5047 Gullen Mall, Detroit, MI 48202

Tel: 3135776708

Fax: 3135776200

Email: mskocmb.biosci.wayne.edu

Seq primer: M13 Reverse.

location/Qualifiers

1..507

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="ATCC (inhost):1553226"

/db_xref="taxon:10090"

/clone="C0004A08"

/clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA

library"

/sex="unknown"

/tissue_type="ectoplacental cone"

/dev_stage="embryonic day 7.5 postconception"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pSPORT1 (Life Technologies);

Site 1: SalI, Site 2: NotI. Total RNAs were extracted from

ectoplacental cone of 7.5-dpc embryos. The

double-stranded cDNA was synthesized from total RNAs with

an Oligo(dT) primer. The library was constructed by Minoru

S. H. Ko."

BASE COUNT 145 a 98 c 106 g 156 t 2 others

ORIGIN

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Percent identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x AA407566

Align seq 1/1 to: AA407566 from: 1 to: 507

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuAqAspSerGly 17

|||||
12 TTCTACITTAACAAACCAAGGCTTCTATGCTAGGAGAAATGACCA 61

17 SASPARGCysicuhisGluSerIysGlyArg 27
|||||
62 GCACGCTGCTTACACGAGTCAAAAAGGCTGG 92

seq_name: qb_est1:AA041885

seq_documentation_block:

LOCUS AA041885 508 bp mRNA linear EST 03-SEP-1996
DEFINITION mJ09g11.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:475268 5' similar to FIP A49733 A49734
glycosaminoglycan N acetylglucosaminyl N deacetylase , mRNA
sequence.

ACCESSION AA041885

VERSION AA041885.1 GI:1520040

WORDS EST

ORIGIN house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 508)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Theising, H., Wylic, T., Lennon, G., Soares, R., Wilson, R. and
Waterston, K.

TITLE The WashU-HMI Mouse EST Project

UNPUBLISHED (1996)

CONTACT: Marra M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL : contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:286312

Seq primer: 28M13 rev2 from Amersham

High quality sequence stop: 490.

FEATURES

source

1. .508
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:475268"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10H"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TGTTACCACTGAAGTGGAGCGCCGCCGAAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT

ORIGIN

136 a 134 c 114 q 124 t

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Length: 27

Gaps: 0

alignment_block:

US-09-557-262-2_copy_250_276 x AA041885

Align seg 1/1 to: AA041885 from: 1 to: 508

1 PheTyrPheAsnLysThrIysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
196 TTCTACITTAACAAACCAAGGCTTCTACTGCTCTGGACAGTGTGAA 245

17 SASPARGCysicuhisGluSerIysGlyArg 27

|||||

246 GCACTGCTGCTTACACGAGTCAAAAAGGCTGG 276

seq_name: qb_est2:BE851982

seq_documentation_block:

LOCUS BE851982 555 bp mRNA linear EST 26-SEP-2000
DEFINITION uw12e07.x1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:3416484
3' similar to TP-045410 045410 HEPATAN SULFATE GLUCOSAMINYL
3-O-SULFOHANSFERASE PRECURSOR ; mRNA sequence.

ACCESSION BE851982

VERSION BE851982.1 GI:10310321

WORDS EST

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 555)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index

UNPUBLISHED (1997)

Other ESTs: uw12e07.y1

Contact: Robert Strausberg, Ph.D.

Email: cqpbs@mail.nih.gov

This clone is available royalty-free through LLNL : contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1092296

High quality sequence stop: 474.

FEATURES

source

1. .555
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3416484"
/clone_lib="Soares mouse 3NBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10H"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TGTTACCACTGAAGTGGAGCGCCGCCGAAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT

ORIGIN

173 a 111 c 114 q 157 t

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Length: 27

Gaps: 0

alignment_block:

US-09-557-262-2_copy_250_276 x BE851982/rev

TECHNICAL ASSISTANT

17 sAspArqCysLeuHisGluSerLysGlyArg 27
|||||
438 GGACCGCTUCTTACACGACTCCAAAGGCCCG 408


```

VERSION      A1870763.1  GI:5544731
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS      1 (bases 1 to 604)
TITLE        NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
              National Cancer Institute / National Institute of Neurological
              Disorders and Stroke, Brain Tumor Genome Anatomy Project
              (CGAP/BTGP), Tumor Gene Index
JOURNAL      Unpublished (1998)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D.,
              Ph.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
              Bonaldo, Ph.D.
              cDNA Library Arrayed by: Greg Isomura, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone Distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              www.bio.llnl.gov/bbrp/image/image.html
              Insert length: 1024  Seq Error: 0 00
              Seq primer: -40mp from Gibco
              High quality sequence stop: 451.
FEATURES     Location/Qualifiers
              1..609
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="IMAGE:2430841"
               /clone_lib="NCI_CGAP_Brn25"
               /tissue_type="anaplastic oligodendrolioma"
               /lab_host="pH10B"
               /note="organ: brain; Vector: pTT730-Pac (Pharmacia) with a
               modified polylinker, Site_1, Not I, Site_2, Eco RI; 1st
               strand cDNA was primed with a Not I - oligo(dT) primer 15'
               TGTACCAATCGAAGCGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTTTT
               T 3'; double-stranded cDNA was ligated to Eco RI
               adaptors (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of the modified pTT73 vector
               library is normalized, and was constructed by Bento
               Soares and M. Fatima Bonaldo."
              HASE COUNT 144 a 201 c 148 g 115 t 1 others
              ORIGIN
              alignment_scores:
                Quality: 1024.00      length: 202
                Ratio: 5.172          Gaps: 0
                Percent Similarity: 98.020  Percent Identity: 90.594
              alignment_block:
                US-09-557-262_2_COPY_21_307 x A1870763
                Align seq 1/1 to: A1870763 from: 1 to: 609
                75 AspTrpLeuGluHisTyrSerGlnGlyLeuGlyTyrPheLeuThrGlnMe 91
                   |||||
                   1 GACTCGGAGGAGCATACAGCCAGCGCTTCGGTTCGTACCTCAGCCAGAT 50
                91 tProPheSerSerProHisGlnLeuThrValGluTysThrProAlaTyrP 108
                   |||||
                   51 GCGTTTCTGAGGAGAAAGATGAGATGAGATGAGATGAGATGAGATGAGAT 100
                   |||||
                   108 heThrSerProGlyValProGluArgGlnHisSerMetAsnProThrIle 124
                   |||||
                   101 TCACCTGCGCCCAAGATGCGGAGCGAGAGATACAGCAIAGAACCGGTCATC 150
                   |||||
                   125 ArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 141
                   |||||
                   151 GCGTTCGTCTCATCTCTGCGAGAGCGGTCGAGAGCGGTCGATCTGATCTA 200

```

```

141 rThiClnValLeuTyrAsnHisLeuGlnGlyHisTyrPheTyrProProI 158
   |||||
201 CACCAACATGTTTATACAACTACATGAGAAATAAATGATACGCGCA 250
   |||||
158 LeuGluAspLeuLeuMetAlaAspGlyArgLeuAsnLeuAspTyrTyrAla 174
   |||||
251 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
   |||||
175 LeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPhePh 191
   |||||
401 GTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
   |||||
191 eProLeuGlyHisIleHisIleValAspGlyAspArgLeuLeuArgAsp 208
   |||||
351 CCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 400
   |||||
208 ProPheProGlnIleGlnValGluAlaPheLeuLeuLysLeuSerProGln 224
   |||||
401 CTTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
   |||||
225 IleAsnAlaSerAspPheTyrPheAsnLysThrLysGlyPheTyrCysLe 241
   |||||
451 ATCAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
   |||||
241 GATGATSerClyLysAspArgCysLeuHisGlnSerLysClyArgAlaH 258
   |||||
501 GCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 550
   |||||
258 IsProGlnValAspProCysLeuLeuLeuAspLysLeuHisGluTyrPheHis 274
   |||||
551 ACCCGCAAGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
   |||||
275 GluPro 276
   |||||
601 GAGCCA 606
seq_name: gb_est2:BC820479
seq_documentation_block:
LOCUS      BC820479              768 bp      mRNA      linear      2001
DEFINITION 502782294F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933676
ACCESSION  BC820479
VERSION    BC820479.1  GI:14168066
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS     NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
JOURNAL     Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: David N. Louis, M.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LLAM10861 row: q column: 21
              High quality sequence stop: 777.
FEATURES     Location/Qualifiers
              1..788
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="IMAGE:4933676"
               /clone_lib="NCI_CGAP_Brn67"
               /tissue_type="anaplastic oligodendrolioma with 1p/19q
               loss"
               /lab_host="DH10B (11 phage-resistant)"

```


/note="Organ: brain; Vector: pCMV-Sport6; Source: NCI-Net; Site: 2; Salt: Cloned unidirectionally; Primer: oligo-dT; Average insert size: 2.3 kb; Constructed by Life Technologies. Note: this is a NCI_CGAP library"

BASE COUNT 209 a 224 c 174 g 181 t

ORIGIN

alignment_scores:

Quality: 1019.00 Length: 202
Ratio: 5.060 Gaps: 0
Percent Similarity: 98.020 Percent Identity: 91.089

alignment_block:

US-09-557-262-2_COPY_21_307 x BG820479 ..

Align seq 1/1 to: BG820479 from: 1 to: 788

```

86 TrpTyrLeuThrGlnMetProPheSerProHisGlnLeuThrValG1 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 TGGTACCTCAGCCAGATGCGCTTCTCTGGGCAACAGCTCAAGTGA 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 uLysThrProAlaTyrPheThrSerProLysValProGluAroLleHis 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 GAAGACCCCGCGTATTTCACGTCGCCCAAGAGUCCCHACCGAGCTACA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 erMetAsnProThrIleArgLeuLeuLeuLeuLeuArgAspProSerGlu 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 GCATCAACCGCTCATCGATGCTGCTGCTATCTGAGAGAGCTGAG; 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 ArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH1 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 GAGGTGCTATCTGACTAAGCCAACTTTTACAAACACATGAGAGATGA 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 sLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeuA 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 CAAGCCGACCCGACCTCCAGCAGCTTCCCTCCGCGATCCCGAGCTCA 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 sLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLeu 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 ATGTGGACTACAAAGCCCTCAACGCGAGCCTCTACGACGTCACATGAG 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 AsnTyrPheArgPhePheProLeuGlyHisIleHisIleValAspGlyAs 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 AACTGGCTGGCTTTTTCGCCGTCGCGCCACATCCACATTTGTGAGCGGA 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 pArgLeuLeuArgAspProPheProGluIleGlnLysValGluArgPheL 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 CCGCCTCATCAGGAGCCGCTTCCCTGAGATCCCAAAAGCTCGAGAGTTC 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 euLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysThr 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 TAAAGTTPGNGGAGAGATGATGTTGAAATTTTAAATAAAATTA 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 LysGlyPheTyrCysLeuArgAspSerClyLysAspArgCysLeuHisG1 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 AAGGGTTTATTAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 uSerLysGlyArgAlaHisProGlnValAspProLysLeuLeuAspLysL 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 GTCCAAAGCCCGCGCGCCACCCCAAGTCGATCCCAAACTACTCAATAAC 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 euHisClyTyrPheHisGluProAsnLysLysPhePheLysLeuValGly 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
553 TGAATGATATTTTCATAGGCTTAATAAGAAAGTTCTTCGAGCTTGTTGC 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 ArgThr 287
|||||
603 AGAACA 608

```

seq_name: qb_est2:BE676274

seq_documentation_block:

LOCUS

DEFINITION BE676274 1 c1-10036815 mRNA 1000 bp
7f26302 x1 NCI_CGAP_C111 Homo sapiens cDNA clone IMAGE:3295779 3' similar to TP-014792-014792 HEPAPAN SULFATE 3 C SULFOTRANSFERASE-1
PREVIOUSOR: 1; mRNA sequence.

ACCESSION

VERSION BE676274 1 c1-10036815

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eumetaria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 621)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Hyrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: infoimage.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 383.

FEATURES

SOURCE

1..621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3295779"
/clone_lib="NCI_CGAP_C111"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="pH108"
/note="Vector: pT73D Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', 10TAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Falima Bonaldo."

BASE COUNT 141 a 204 c 150 g 125 t

ORIGIN

alignment_scores:

Quality: 1007.00 Length: 207
Ratio: 5.060 Gaps: 0
Percent Similarity: 96.145 Percent Identity: 86.957

alignment_block:

US-09-557-262-2_COPY_21_307 x BE676274 ..

Align seq 1/1 to: BE676274 from: 1 to: 621

```

71 ValHisPhePheAspTyrGluGluHisTyrSerGlnGlyLeuGlyTrrpy 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 GTCCACTTCTTCGACTGGGAGGAGCATTTACAGCCACCGCTTGGCGTGA 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 rLeuThrGlnMetProPheSerSerProHisGlnLeuThrValHisLysT 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 CCTCAGCCAGATGCGCTTCTCTGCGCCACACACACCTCACACTGCGAGA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 hrProAlaTyrPheThrSerProLysValProGluArgLleHisSerMet 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 CCCCCTGCGTATTTCACGTCGCCCAAGAGGCTTAAGAGGTAAAGATG 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 AsnProThrIleArgLeuLeuLeuLeuLeuArgAspProSerGluArgVa 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

151 AACCGGTGATGAGTGTCTGTTATGCTGAGAAAGAGGAGAGAGT 200
157 lleuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnIysHisLysP 154
|||||
201 GATATGTAATATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 250
154 roTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeuAsuLeu 170
|||||
251 CCAACCGGTCATGAGAGAGTCTGGTGGGAGGAGGAGGAGGAGGAGG 300
171 AspTyrIysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTr 187
|||||
301 GACTATCAAGAGCTTCAAGCGTAAAGCTTATCAAGATGTAATGTAAG 350
187 pleuArgPhePheProIleGluGlyHisIleHisIleValAspGlyAspArg 204
|||||
351 GCTGGCGGTTTCTGGGATGGAGTCAATGACATGACATGAGGAGGAGG 400
204 eufLeuArgAspProPheProIleGlnIysValIleuArgPheLeuLys 220
|||||
401 TCATCAGGAGTCTCTCCCTGCACATCAAAAGCTCGAGAGGTCTCTAATC 450
221 leuSerTrpGlnIleAsnAlaSerAsnPheTyrPheAsnLysThrLysG 237
|||||
451 CTGTGGCGGTAAGATCAATGATCTGAAATTTTAAATTTTAAATTTTAA 500
237 yPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHisGluSerL 254
|||||
501 CTTTATTCTGCTGCTGCACACGCGCGCGGACCGTCTCTTACATGACTCA 550
254 ysdGlyArgAlaIleHisProGlnValAsnProLysLeuLeuAspLysLeuHis 270
|||||
551 AATCGCGGCGGACATGACATGACATGACATGACATGACATGACATGAC 600
271 GluTyrPheHisGluProAsn 277
|||||
601 GAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 521

```

seq_name: gb_est2_HCR20537

```

seq_documentation_block:
LOCUS      BCR20537              783 bp     mRNA     linear     EST 22-MAY-2001
DEFINITION 602782876F1 NCI_CGAP_Brc67 Homo sapiens cDNA clone IMAGE:4933675
5' mRNA sequence.

```

ACCESSION HCR20537.1 GI:14168124

VERSION HCR20537.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates, Catarrhini, Hominoidea, Homo.

1 (bases 1 to 783)

NIH-MGC http://mgc.ncl.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph D

Email: cgapbs@mail.nih.gov

Tissue procurement: David N. Louis, M.D.

cDNA library preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: LAM10461 row: g column: 20

High quality sequence stop: 783.

alignment_scores:

Quality: 1006.00

Ratio: 5.081

Percent Similarity: 97.537

Percent Identity: 90.640

alignment_block:

US-09-557-262-2_copy_21_307 x HCR20537

Align seg 1/1 to: BCR20537 from: 1 to: 783

86 TrpTyrLeuThrGlnMetProPheSerProHisGlnLeuThrValG 102

2 GGTATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 51

102 uLysThrProAlaTyrPheThrSerProLysValProGluArgIleHis 118

52 GAATATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 101

119 SerMetAsnProThrIleArgIleuLeuLeuLeuLeuLeuLeuLeu 135

102 AATGATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 151

135 uArgValIleuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnIysH 152

152 GCGGCTGCTGCTGCACACGCGCGGACCGTCTCTTACATGACTCA 201

152 isLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeu 168

202 ACAAGATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 251

169 AsnLeuAspTyrIysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185

252 AATGATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 301

185 uAsnTrpPheLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 202

302 GAATGCTGCGGCTTTTCTCCCTGCACATCAAAAGCTCGAGAGGTCT 351

202 spArgLeuIleArgAspProPheProGlnIleGlnIysValGluArgPhe 218

352 AATGATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 401

219 LeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysTh 235

402 GTAAATCTGATGAGTCAAGTGGTGTATCAAAATGATGACAAACAAAG 451

235 rLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHisC 252

452 CAAGTGGCTTTTATGAGTGGGAGACACAGCGGCGGACCGCTCTTACATG 501

252 luSerLysGlyArgAlaIleHisProGlnValAspProLysLeuLeuAspLys 268

502 AGTCCAAAGGCGGCGGAGACAGCGGCGGAGTGGTGGTGGTGGTGGTGG 551

269 LeuHisGluTyrThrHisGluProGlnLysIleHisIleHisIleHisIle 285

552 CTCACAGCAATATTTTCATGAGCCCAATATGAGTCTCTTCTGAGCTTCT 601

285 YArgThr 287

602 CAGAACA 608

seq_name: gb_est1:AI582352

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4933675"

/clone_lib="NCI_CGAP_Brc67"

/tissue_type="anaplastic oligodendroglioma with 1p/19q

```

seq_documentation_block.
LOCUS      AF582352              618 bp      mRNA      linear      EST 14-DEC-1999
DEFINITION      lq67b01.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2214844.3'
                similar to TP-014792 HEPAPAN SULFATE 3'-UTR-5'UTR-TRANSPASE-1
                PRECURSOR : mRNA sequence.
ACCESSION      AF582352
VERSION        AF582352.1 GI:4568249
KEYWORDS       EST.
SOURCE         human
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
                Emmert-Buck, M.D., Ph.D.
                cDNA Library Prepared by: M. Bento Soares, Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www-bio.llnl.gov/dbbrp/image/image.html
                Insert Length: 930 Std Error: 0.00
                Seq primer: -400p from GIBCO
                High quality sequence stop: 456
                POLYA-No.
FEATURES             Location/Qualifiers
     source           1. 618
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2213833"
                     /clone_lib="NCI-CGAP_Lu19"
                     /tissue_type="squamous cell carcinoma, poorly
                     differentiated (4 pooled tumors, including primary and
                     metastatic)"
                     /dev_stage="adult"
                     /lab_host="DH10b (phage-resistant)"
                     /note="Organ: lung; Vector: pTZ19-lac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     pooled lung tumor tissue, and was then primed with a Not I
                     - oligo(dT) primer. Double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pTZ3 vector. Library went through one round of
                     normalization. Library constructed by Bento Soares and M
                     Fatima Bonaudo."
     BASE COUNT       144 a 201 c 149 g 119 t 5 others
     ORIGIN
alignment_scores:    Quality: 986.00      Length: 206
                     Fail: 5.056         Gaps: 0
Percent Similarity: 94.650 Percent Identity: 86.408
alignment_block:
US-09-557-262-2_COPY_21_307 x AF582352
Align seg 1/1 to: AF582352 from: 1 to: 618
75 AspTrpGluGluHisTyrSerGlnGlyLeuGlyTrpTrpLeuThrGlnGme 91
|||||
1 GACTGGAGGAGCATTACAGCAGGCTTG3GCTGGTAACCTACAGCCAGAT 50
91 tProPheSerProHisGlnLeuThrValGluGlyThrProAlaTyrP 108
|||||
51 GCCCTCTCTCGCCGACACACAGCTCAGAGTGGAGAGACCCCGCGTAVT 100

```

```

108 heHsrSerProLysValProGluArgIleHisSerMetAsnProThrIle 124
|||||
101 TCACATGCGGAAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 150
|||||
125 ArgLeuLeuLeuLeuLeuArgAspProSerGluA: gValLeuSerAspTy 141
|||||
151 CGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 200
|||||
141 rThrGluValLeuTyrAsnHisLeuGlnLysHisLysProTyrProProI 158
|||||
201 CACCAAGCTGTGTACCAACACACACACACACACACACACACACACACAC 250
|||||
158 leGluAspLeuLeuMetArgAspGlyArgLeuAsnLeuAspTyrLysAla 174
|||||
251 TGAAGAAATTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 400
|||||
175 LeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPhePh 191
|||||
301 CACAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 350
|||||
191 eProLeuGlyHisIleHisIleValAspGlyAspArgLeuLeuIleArgAsp 208
|||||
351 GCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400
|||||
208 rPheTrpGluGlnGlnLysValGluArgPheLeuLysLeuSerTrpGln 224
|||||
401 CTTTGTGTGATATGAAAGGTGATGATGATGATGATGATGATGATGATG 450
|||||
225 lLeAsnAlaSerAsnProTyrPheAsnTyrHisLysGlyCysCysCys 241
|||||
451 ATCAATGCTTCCAACTCTACTACTACTACTACTACTACTACTACTACTACT 500
|||||
241 uArgAspSerGlyLysAspArgCysLeuHisGlnSerLysLysValArgAla 258
|||||
501 GCGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
|||||
258 isProGlnValAspProLysLeuLeuAspLysLeuHisGluTyrPheHis 274
|||||
551 ACCGCTAGTGTGATATGAAAGGTGATGATGATGATGATGATGATGATG 600
|||||
275 GluProAsnLysLysPhe 280
|||||
601 AAGCCAAAGAGAGCTCTTC 618
seq_name: qb_est2:BF447174
seq_documentation_block:
LOCUS      BF447174              612 bp      mRNA      linear      EST 01-DEC-2000
DEFINITION      7p45u11.x1 NCI-CGAP_P128 Homo sapiens cDNA clone IMAGE:3648861.3'
                similar to TP-014792 HEPAPAN SULFATE 3'-UTR-5'UTR-TRANSPASE-1
                PRECURSOR : mRNA sequence.
ACCESSION      BF447174
VERSION        BF447174.1 GI:11512312
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
                Emmert-Buck, M.D., Ph.D.
                cDNA Library Prepared by: M. Bento Soares, Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:

```

info@image.llnl.gov
 Seq primer: -400p from Gibco
 High quality sequence stop: 474
 Location/Qualifiers
 1. 612
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:364861"
 /clone_lib="NCL_CGAP_Pr28"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ. prostate. Vector. p713b-Pac (Pharmacia) with a modified polylinker. Plasmid DNA from the normalized library NCL_CGAP_Pr28 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Benito Soares and M. Fatima Bernaldo."

BASE COUNT 138 a 202 c 150 g 121 t 1 others
 ORIGIN

alignment_scores:

Quality: 981.50 Length: 214
 Ratio: 5.033 Gaps: 1
 Percent Similarity: 91.121 Percent Identity: 84.112

alignment_block:

US-09-557-262-2_copy_21_307 x BF447174 ..

Align seq 1/1 to: BF447174 from: 1 to: 612

71 ValHisPhePheAspTrpGluGluHisLysSerGlnGlyLeuGlyTrp 87
 1 GGCACATCTTCGACTGGAGGAGCATACAGCCACGGCTTGGCTGGCA 50
 87 rLeuThrGlnMetProPheSerProHisGlnLeuThrValGluLys 104
 51 CCTCAGCCAGATGCTTCCTCTCTGCTGCTGCTGCTGCTGCTGCTG 100
 104 hrProAlaTyrPheThrSerProLysValProGluArgIleHisSerMet 120
 101 GCGGGGCTATTCACGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 150
 121 AsnProThrIleArgIleLeuLeuLeuArgAspProSerGluArgVa 137
 151 AACCCGTCCTCACTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 200
 137 IleuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysHisLysP 154
 201 GCTATCTGACTACACCAAGCTGTTTACAACTACATGCAAGCAACCA 250
 154 rGlyProProIleGluAspLeuLeuMetArgAspGlyArgGluAsnLeu 170
 251 CTATGCTGCTATGAGAAATTTCTGATGCTGCTGCTGCTGCTGCTG 400
 171 AspTyrLysAlaIleuAsnArgSerLeuTyrHisAlaHisMetIleuAsn 187
 301 CACTACAGGCTTCACCCGACGCTTACCCAGCTGCTGCTGCTGCTGCT 350
 187 pLeuArgPhePheProLeuTyrHisIleHisIleValAspLysAspArg 404
 351 GCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400
 204 cGlyLeuArgAspThrPheProGlnIleGlnLysValIleuArgPheLeuLys 220
 401 TCATCAGGAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 450
 221 LeuSerProGlnIleAsnAlaSerAspPheTyrPheAsnLysThrLysG 237

451 CTGTGCGCGCAGCATCAATGCT..... 471
 237 yPheTyrCysLeuArqAspSerGlyLysAspArqCysLeuHisGluSerL 254
 472 TATGCTGTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
 254 YSGLYARGAlaHisProGlnValAspProLysLeuLeuAspLysLeuHis 270
 518 AANAGATGATGAGACACGCTGAGATGATGATGATGATGATGATGATG 567
 271 GluTyrPheHisGluProAsnLysLysPhePheLysLeuVal 284
 568 CAATATGATGATGAGACACGCTGAGATGATGATGATGATGATGATG 609

seq_name: qb_est2:BG746348

seq_documentation_block:

LOCUS BG746348 783 bp mRNA linear EST 15-MAY 2001
 DEFINITION 602703607F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857025 5',
 mRNA sequence.

ACCESSION BG746348 1 GI:14057001
 VERSION BG746348
 KEYWORDS EST.

SOURCE

human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 783)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov

Tissue procurement: ATCC

CDNA library Preparation: Ling Hong/Kubin Laboratory

CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1708 row: n column: 02

High quality sequence stop: 720.

FEATURES

location/Qualifiers
 1..783
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4857025"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pUTB7; Site: 1; XhoI; Site: 2;
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(C). Size-selected, 500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 181 a 255 c 176 g 171 t
 ORIGIN

alignment_scores:

Quality: 944.50 Length: 222
 Ratio: 4.563 Gaps: 7
 Percent Similarity: 93.243 Percent Identity: 83.784

alignment_block:

US-09-557-262-2_copy_21_407 x BG746348 ..

Align seq 1/1 to: BG746348 from: 1 to: 783

73 PhePheAspTrpGluGluHisTyrSerGlnGlyLeuGlyTrpLysLeu 89

REFERENCE

1 (bases 1 to 562)
 NC/NIHNS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BRGAP), Tumor Gene Index

JOURNAL

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: rga@rslmail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D.,
 Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCICGAP clone distribution information can be
 found through the I M A G E Consortium/LLNL, send email to:
 info@image.llnl.gov

Seq primer: -40up from Gibco

High quality sequence stop: 370.

Location/Qualifiers

FEATURES

source

1..562

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAG:3563691"

/clone_lib="NCICGAP_Brn23"

/tissue_type="glioblastoma (pooled)"

/lab_host="DH10B"

/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TCGTACCAATCGACGCGAGCGGCCGCAATCTTTTTTTTTTTTTTTTTTTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT7T3 vector.
 Library is normalized, and was constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 120 a 185 c 133 g 123 t 1 others

ORIGIN

alignment_scores:

Quality: 865.00 Length: 186

Ratio: 4.971 Gaps: 0

Percent Similarity: 93.548 Percent Identity: 84.409

alignment_block:

US-09-557-262-2_COPY_21_307 x BF594148 ..

Align seg 1/1 to: BF594148 from: 1 to: 562

70 GluValHisPhePheAspTrpGluGluHisTyrSerGlnGlyLeuGlyTr 86

|||||

3 GAAGTCCAACTTCTTCGAATGGAGAGATTATACAGCCAGCGCTGGGCTG 52

86 ptyrLeuThrGlnMetProPheSerProHisGlnLeuThrValGluL 103

|||||

53 GTTCTCAGCCAGATGCGCTTCTCTGCGCACACCCAGCTCACAGTGGAGA 102

103 ysThrProAlaTyrPheThrSerProLysValProGluArqIleHisSer 119

|||||

103 AGACCCCGCGGTTTTCACGTGCGCCAAAGTGCGCTGACCGAGTCTACAGC 152

120 MetAsnProThrIleArqLeuLeuIleLeuArqAspProSerGluAr 136

|||||

153 ATGAACCCGTCATCGGCTGCTGCTCATCTGGAGACCGCTCGGAGCG 202

136 qValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysHisL 153

|||||

203 CGTGTATCTGATACACCCAGGTTTCTTACNACCACATGCGAGAGCACA 252

153 ysProTyrProProIleGluAspLeuLeuMetArqAspGlyArqLeuAsn 169

|||||

253 AGCCCTACCCGTCGATGAGGAGTTCTCTGCTGGCGGAGTGGCGCTCAAT 302

170 LeuAspTyrTyrLysAlaIleuAscArqSerLeuTyrHisAlaHisMetLeuAs 186

|||||

303 GAGAACATACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 352

186 nTrpLeuArqPhePheProLeuGlyHisIleHisIleValAspGlyAspA 203

|||||

363 CAGGCTGCGGCTTTTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402

203 rqlLeuIleArqAspProPheProGluIleGlnLysValGluArqPheLeu 219

|||||

403 GCGTCATCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 452

220 LysLeuSerTrpTrpGlnIleAsnAlaSerAsnPhcTyrHisAsnLysThrL 236

|||||

453 AAGTGTGTGGGCGGAGTAAATGAGCTTTTAAATTTTAAATTTTAAATTTT 502

236 sGlyPheTyrCysLeuArqAspSerGlyLysAspArqCysLeuHisGlnL 253

|||||

503 AGGCTTTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552

253 erLysGly 255

|||||

553 CCAAGGCG 560

Align seq 1/1 to: AAX37245 from: 1 to: 1685

```
alignment_block:
```

US-09-557-262-2 COPY 24 397 X AAX37246

Align seq 1/1 to: AAX37246 from: 1 to: 1305

```

2  ProAlaIaProGlyProGlyLeuLysGlnGlnLeuLeuArgLysVa 18
   |||||
173 CTTTCCTCCCGCCCGCCGAG CTAGCCACAGCAGAGCTTCTGCGGAAAGC 219
   |||||
18  IfleleleuProGluAspThrGlyGlyGlyThrAlaSerAsnGlySer 35
   |||||
220 GGGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 264
   |||||
35  hrGlnGlnLeuProGlnThrIleIleIleIleClyValArgGlyGlyThr 51
   |||||
270 CCGACAGCTTCCCGCAGACCATCATCATCGCCCTCCCGCAAGCGCGCAG 319
   |||||
52  ArgAlaLeuLeuGluMetLeuSerLeuLysProAspValAlaAlaGln 68
   |||||
320 CCGCCAGCTTCCCGCAGACCATCATCATCGCCCTCCCGCAAGCGCGCAG 369
   |||||
68  uAsnGluValHisPheAspThrGlyGlyGlyHisTyrSerGlnGlyLeu 85
   |||||
370 GAAAGAGGTCCATTTCTGAGTGGAGAGGATTAATAGGAGAGGTTGG 419
   |||||
85  lyTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrVal 101
   |||||
420 GCTGTACCTCAGCCAGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 469
   |||||
102 GluLysThrProAlaTyrPheSerProLysValProGluArgIleHis 118
   |||||
470 GAGAAGACCCCGCGTATTTCCAGCTCCCGCAAGCTGCTGAGCAGCTTA 519
   |||||
118 sSerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProSerG 135
   |||||
520 CAGCATGAAGAGGTCATCTGAGGATCTGCTGAGGATCTGAGGATCTG 569
   |||||
135 luArgValLeuSerAspThrThrGlnValLeuTyrAsnHisLeuGlnLys 151
   |||||
570 ACCGCGCTGCTATCTGATACACCAAGTGTTTACACCAACCATGCGAAG 619
   |||||
152 HisLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLe 168
   |||||
620 CACAAGCCCTACCGCTCCATCCAGGAGCTTCTTCTTCTTCTTCTTCTT 669
   |||||
168 uAsnLeuAspTyrLysAlaLeuAsArgSerLeuTyrHisAlaHisMet 185
   |||||
670 CAATGTGATACAAAGGAGGATTAATGAGAGGATTAATGAGAGGATTA 719
   |||||
185 euAsnTyrLeuArgPhePheProLeuGlyHisIleHisIleValAspCly 201
   |||||
720 ACAACTCGCTCGCTTTTCCCGCTCCCGCCACATCCACATCTCGACCGC 769
   |||||
202 AspArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 218
   |||||
770 GACCGCGCTCATCAGGAGCCCTTCCCTGAGATCCAAAAGCTTCGAGAGCT 819
   |||||
218 eleuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysT 235
   |||||
820 CTAAAGTGTGAGGAGATTAATGAGAGGATTAATGAGAGGATTAATGAG 869
   |||||
235 hrLysClyThrCysLeuArgAspSerGlyLysAspArgCysLeuHis 251
   |||||
870 CCAAGCGCTTTTACTGCTTGGCGGACAGCGCGCGGACCGCTGCTTACAI 919
   |||||
252 GluSerLysGlyArgAlaLysProGlnValAspProLysLeuLeuAspLys 268
   |||||
920 GAGTCCAAAGCGCGCGGCTACCGCCCAAGTCCATCCCAAACTACTCAATA 969
   |||||
268 sSerHisGluTyrPheHisGluProAsnLysLysPhePheLysLeuValG 285
   |||||
970 ACTGACGAATATTTTATGAGATTAATAGAGGATTTTGAAGTTGTTT 1019
   |||||
285 lyArgThr 287
   |||||

```

1020 GCAGAAC 1027

seq_name: /SDS5/qcodata/geneseq/geneseq-emb1/NA2000.DAT:AAZ36687

seq_documentation_block:

ID AAZ36687 standard; DNA; 1305 BP.

XX AAZ36687;

XX 14 MAP-2000 (first entry)

XX Nucleotide sequence for GenBank accession number AF019386.

XX Stimulus-regulated nucleic acid; sequence profile; nucleic acid level;
 XX differentially expressed nucleic acid; disease state; cancer;
 KW autoimmune disease; infectious disease; aging; developmental disorder;
 KW proliferative disorder; neurological disorder; toxicity;
 KW treatment resistance; differential expression; drug discovery;
 KW growth factor; epidermal growth factor; radiation; stress; pathogen; ss.

XX Homo sapiens.

XX WO955913-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-0509119.

XX 27-APR-1999; 98US-0084331.

XX 27-AUG-1998; 98US-0098070.

XX 04-FEB-1999; 99US-0118624.

XX (KIMM-) KIMMEL CANCER CENT. SIDNEY.

XX McClelland M, Welsh J, Trenkle T;

XX WPI; 2000-086388/07.

XX Measuring expression of low abundance reduced complexity target nucleic acid molecules -

XX Disclosure: Fig 14; 137pp; English.

XX AAZ36687-230725 represent nucleotide sequences from stimulus regulated nucleic acid molecules. The sequences represent a profile of sequences which can function as targets in the method of the invention. The specification describes a method for measuring the level of two or more nucleic acid molecules in a target. The method comprises contacting a probe with an arbitrarily or statistically sampled target and detecting the amount of specific binding of the target to the probe. The methods can be used to identify differentially expressed nucleic acid molecules associated with disease states, such as cancer, autoimmune disease, infectious disease, aging, developmental disorder, proliferative disorder or neurological disorder. Alternatively the methods can be used to assess the efficacy of toxicity of or a resistance to a treatment. Also the methods can be used to determine differential expression of nucleic acid molecules in response to a stimulus, e.g. a chemical, drug or growth factor (especially epidermal growth factor), radiation, stress or a pathogen. The methods can also be used to determine over-regulated genes that can be potential targets for drug discovery.

XX Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:

Quality: 1345.50 Length: 286

Ratio: 4.929

Percent Similarity: 95.455 Percent Identity: 87.413

alignment_block:

US-09-557 262 2_COPY_21_307 x AAZ36687

Align seq 1/1 to: AAZ36687 from: 1 to: 1305

2 ProAlaIaProGlyProGlyLeuPheLysGlnGlnGluLeuLeuArgLysVa 18
173 GCTTACAGGGGATGAGGAGGTTAAAGGAGAGAGATTGTTCAGAAAAA 219
18 TTTTtTGleuProGluAspThrGlyGluGlyThrAlaSerAsnGlySer 35
220 GGAGAGGTTTAAGAATAAGAGAGGATGAATGAGTGGGAAAAGATTT 269
35 ArgGlnGlnProGlnThrIleIleLeGlyValArgLysGlyGlyThr 51
270 CCAACAATGCTGGCGGCACAAATCAATACAGCGGAGGAGGAAAGGACG 319
52 ArqAlaLeuLeuGluMetLeuSerLeuLisProAspValAlaAlaGla 68
320 CGAGACTTGAATGAATATTAAGAGTAGATGATGATGATGATGATGAT 469
68 uAsnGluValHisPhePheCaspTrpGluGluHisIstYserGlnGlyLeuG 85
370 GANAAGAGTAAATTTTTCATGATGAGAGAGAAATTAAGAACAGGATTGG 419
85 TyrTrpTyLeuThrGlnMetProPheSerSerProHisGlnLeuThrVal 101
420 GCTGCTGATGCTAGCATAGAGGATGATGATGATGATGATGATGATGAT 469
102 GluLysThrProAlaTyrPheThrSerProLysValProGluArqGlyHis 118
470 GANAAATGATGAGGATATTAAATGAGGAAAATGATTAAGATGATTTT 519
118 sSerMetAsnProThrIleArgGluLeuLeuIleCuArqAspProSerG 135
520 CAGCATGAACGGTGCATGAGATGATGATGATGATGATGATGATGATGAG 569
145 LuArqValLeuSerAspTyrThrGlnValLeuTyrAsnLisLeuGlnLys 151
570 ACGCGCTGCTATCTGCATCACCCCAAGTGTTCTACCAACCACATGCCAAG 619
152 HisLysProTyrProProIleGluAspLeuLeuMetArqAspGlyArqLe 168
620 CACAAAGGATACAGGATGATGAGAGATTTGATGATGATGATGATGATGAT 669
168 uAsnLeuAspTyrLysAlaIleuAsnArgSerLeuLysHisAlaHisMet 185
670 CAATTGTGATTAAGAGGATTAAGAGGATTAAGAGGATTAAGAGGATTAAG 719
185 eAsnTrpLeuArqPhePheProLeuGlyHisIleHisIleValAspGly 201
720 ACAACCTGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATG 769
202 AspArgLeuIleLeuArgAspProPheTrpGluLeuLeuValGluArqGly 218
770 GAAGGATTAATAGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 819
218 eGlySerSerProGlnThrAsnArgAspValTyrPheAspGlySer 235
820 CCTAAGCTGTGCGGCGAGATCAAGGATTCAGAAATTCATATTATTAACAAA 869
245 hrLysGlyPheTyrGlySerLeuArgAspSerGlyLysAspArgGlyLeuHis 251
870 CCACGGCTTTTTACTGCTCGCGGCACACCGCCCGGACCGCTGCTTACAT 919
252 GluSerLysGlyArqAlaIleProHisValAspProLysLeuPheAspLys 268
920 GAGTCCAAAGGCGCGAGGATGATGATGATGATGATGATGATGATGATGATG 969
269 sLeuHisGlnTyrPheHisGlnProAsnGlyLysPhePheHisLysValG 285
970 ACTGACCAATATTTTTATGAGTCAATATGAAATGAAATGATTTTGAATG 1019
285 LysArgThr 287
1020 GCAGAACA 1027

seq_name: /SIDS5/qcadata/qeneseq/qeneseqn-emb1/NA2001A.DAT:AA50047

seq_documentation_block:

ID ABA50047 standard; DNA; 924 BP.

AC AHA5U047:

XX 01-FEB-99 (first entry)

XX
DE Hum30 breast cell cycle nucleic acid probe #0749

XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.

AA
QS
Homo sapiens

XX
PN
W0300157371-A3XX
100-815-3001XX
PF 20-YEN-0001 000100 00000000

XX

DATE MAY - 2000; 2000US 0207456.

99E2830-Suburb - 1000000 - 1111-1111

PP 27-SEP-2000; 200005 11246359

[illegible]XX
XX
(MOL) - MOLECULAR DYNAMICS
XX
XXPI Penn SG, Hanzel DK, Chen W, Rank DR, ^{yy}

DE WPI; 2001-496933/54.

PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -

PS Claim 4; SEQ ID NO 8742; 327bp ; sequence listing: English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct_sequences/.

924 DE VRIES, A., 1993. I: 0 other.

alignment scores.

Quality:	1335, 50	Length:	284
----------	----------	---------	-----

Quantity: 1532.00
Ratio: 4.928
Gaps: 1

Family:	95.423	Percent Identity:	87.324
---------	--------	-------------------	--------

alignment_block;

US-09-557-262 2 COPY 21 307. X ABA50047/TCV


```

754 CCCACACCTCCCGCCACATATATATATCGGCGTGGCGAAGCGCGGACG 705
52 ArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaGlu 68
704 CCGCCACCTCCCGTATATATATATATATATATATATATATATATAT 655
68 uAsnGluValHisPheAspThrGluGluHisTyrSerGlnCylLeuG 85
654 GAAACAGATATATATATATATATATATATATATATATATATATAT 605
85 TyrPtyrLeuThrGluMetProPheSerSerProHisGlnLeuVal 101
604 CCGTGTATATATATATATATATATATATATATATATATATATAT 555
102 GluLysThrProAlaTyrPheThrSerProLysValProGluArgIleHis 118
554 CAGACAGCCCGCGTATATATATATATATATATATATATATATATAT 505
118 sSerMetAsnProThrIleArgGluLeuLeuIleLeuArgAspProSerG 135
504 CAGCATATATATATATATATATATATATATATATATATATATAT 455
135 IuArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151
454 AGCGGCGTATATATATATATATATATATATATATATATATATAT 405
152 HisLysProTyrProProIleGluAspLeuLeuMetArgAspGluArgLe 168
404 CACAAAGCTTACCGTATATATATATATATATATATATATATATAT 355
168 uAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetHis 185
354 CAATGTGAGTATATATATATATATATATATATATATATATATAT 305
185 euAsnThrPheArgPhePheProLeuGluCysHisIleHisIleValAspCly 201
304 ACAACTGCGTCGCGCTTTTCCCGTCGCCACATCCACATTCGACGCG 255
202 AspArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPh 218
254 CACCCCTTCATCAGCGACCCCTTCCCTGCATCCAAAGCTCCAGAGCT 205
218 GluLysLysSerThrProGluIleAspAlaSerAsnProPheTyrPheAsnLys 235
204 CATTAAATATATATATATATATATATATATATATATATATATAT 155
235 hTrpGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHis 251
154 CCAAGCGTATATATATATATATATATATATATATATATATATAT 105
252 GluSerLysGlyArgAlaHisProGluValAspProLysLeuLeuAspLys 268
104 CAGTCCAAAGCGCGCGCGCACCCCAAGTCCATCCCAAACTACTCAATA 55
268 IleuHisCylThrPheHisGluProAsnLysLysPhePheLysLeuValC 285
54 ACTGATATATATATATATATATATATATATATATATATATATAT 5
285 Ty 285
4 GC 3

```

seq_name: /SDS5/qcdata/qceneseq/qeneseq-emb1/NA2001A.DAT:ABA35017

seq_documentation_block:

ID ABA35017 standard: IMA: 924 bp

XX

XX

XX

XX

DT

DE

23-JAN-2002 (first entry)

Probe #13483 for gene expression analysis in human heart cell sample

XX Human: gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-PRB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 03-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0643666.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0324263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SC, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human heart

XX Claim 4; SEQ ID NO 13483; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from Wipo at http://wipo.int/pub/published_pat_sequences.

XX Sequence 924 BP, 169 A, 252 C, 310 G, 193 T; 0 other;

alignment_scores:

Quality: 1435.50

Ratio: 4.928

Percent similarity: 95.423

Percent identity: 87.324

alignment_block:

US-09-557-262-2_COPY_21_307 x ABA35017/rcv

Align seq 1/1 to reverse of: ABA35017 from: 1 to: 924

2 ProAlaAlaProGlyProGlyLeuLysGlnGlnGluLeuLeuArgLysVa 18

851 CCGTCCCGCGCGCGGAG CTAGCCACGACGAGCTTCTCCGCAAGC 805

18 IleIleLeuProGluAspThrGlyGluThrAlaSerAsnGlySerT 45

804 GAGGAGATATATATATATATATATATATATATATATATATATAT 755

35 hGlnGlnLeuProGlnThrIleIleIleIleIleIleIleIleIleIle 51

754 GCGAAAT 705

52 ArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaGlu 68

704 CCGGAT 655


```

554 CAGACAGACCCCGGCTATTTCACGTCGCGCCAAATGCTGCTCAGCCAGCTCTA 505
118 sSerMetAsnProThrIleArqLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 135
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
504 CAAATAT:AAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 455
135 IuArqValIleuSerAspIleThrGlnValIleuValIleuValIleuValIleu 151
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 AGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405
152 HisLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLe 168
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 CACAGAGGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 355
168 uAsnLeuAspTyrLysAlaLeuAsnArqSerLeuTyrHisAlaHisMetL 185
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 CAATGAGGACTA:AAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 305
185 cuAsnTrpLeuArqPhePheProLeuGlyHisIleHisIleValAspGly 203
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 AGAAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255
202 AspArqLeuIleArqAspProPheProGluIleGlnLysValGluArqPh 218
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 GACCCGCTCATCAGGGAGCCGCTTCCTGAGAGAGAGAGAGAGAGAGAGAGAG 205
218 eLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysT 235
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 CATAAAATGTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 155
235 hTrpGlyPheTyrCysIleuArqAspSerGlyLysAspArgCysLeuHis 251
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 CCAGAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 105
252 GluSerLysGlyArqAlaHisProGlnValAspProLysLeuLeuAspLy 268
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 GAGTCCAAAGCCCGGCGCCACCCCAAGTCGATCCCAAACTACTCAATAA 55
268 sLeuHisGlyTyrPheHisGluProAsnLysLysPhePheLysLeuValG 285
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 ACTGACAGATATTTTCATGAGCAATATAGAAATGATGATGATGATGATGAT 5
285 ly 285
      :|
      4 GC 3

```

seq_name: /SID55/gcqdala/geneseq/geneseq-emb1/NA2001A.DAT:AAK42122

seq_documentation_block.

AAK42122 standard; DNA; 924 bp.

AAK42122:

06-NOV-2001 (first entry)

Human bone marrow expressed single exon probe SEQ ID NO: 16679.

Human; bone marrow expressed exon; gene expression analysis; probe; microarray, cancer, leukemia, lymphoma, myeloma, ss.

Homo sapiens.

W0200157276-A2.

09-AUG-2001.

30-JAN-2001: 2001WO-0500668

04-FEB-2002: 2002US-0180312.

26-MAY-2000: 2000US-0207456.

30-JUN-2000: 2000US-new0408

03-AUG-2000: 2000US-0632366.

21-SEP-2000: 2000US-0234687.

```

PR 27-SEP-2000: 2000US-0236359.
PR 04-OCT-2000: 2000GR 0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DK;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow.
XX
XX Example 4: SEQ ID NO: 16679; 658bp. Sequence listing, English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other;
SQ

```

alignment_scores:

Quality: 1335.50 Length: 284
 Pct. Id: 4.928 Gaps: 1
 Percent Similarity: 95.423 Percent Identity: 87.324

alignment_block:

US-09-557 262 2_COPY_21_307 x AAK42122/rev ..

Align seq 1/1 to reverse of: AAK42122 from: 1 to: 924

```

2 ProAlaIaProThrProGlyLeuLysGlnGlnLeuLeuLeuLeuVal 18
851 GCTTCCCGCCCGCCCGGAG...CTAGGCGCAGCAGGAGCTTCTGGCGAAG 805
18 IileIleLeuProGluAspThrGlyGlnGlyThrAlaSerAsnGlySerT 35
804 GGGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 755
35 GlnGlnLeuProGlnThrIleIleIleGlyValArqLysGlyGlyThr 51
754 GCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705
52 ArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaAlaG 68
704 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 655
68 uAsnGlnValHisPhePheAspTrpGlnGlnHisTyrSerGlnGlyLeuG 85
654 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
85 LYTTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrVal 101
604 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
102 GluLysThrProAlaTyrPheThrSerProLysValProGluArqIleu 118
554 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 505
118 sSerMetAsnProThrIleArqLeuLeuLeuLeuLeuLeuLeuLeuLeu 145
504 CAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 455
135 LeuArgValIleuSerAspTyrThrGlnValIleuTyrAsnHisLeuGlnLys 151
454 AGCAGGAGGATATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
152 HisLysProTyrTrpGlnGlnGlnGluLeuLeuLeuLeuLeuLeuLeuLeu 168
404 CATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355

```



```

168 uasnleuasptryltsalaleuasnargserleuTyHisAlaHisMetL 185
|||||
354 CAATGTGCAATACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
|||||
185 euasntrpleuargphepheleuGlyHisIleHisIleVal:AspGly 201
|||||
304 AGAACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 355
|||||
202 AspAlaLeuLeuArgAspProGlyLeuGlnLysValGluArgPhe 218
|||||
254 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 305
|||||
218 cLeuLysLeuSerProGlnLysAlaSerAsnPhelThrAsnLys 235
|||||
204 CCAAGAGCTGGCGCGAGATCAAGCTTGGAACTTCTACTTTAACAAA 255
|||||
235 hrlsclpghelrlycysleuargaspserGlyLysAspArgCysLeuHis 251
|||||
154 CCAAGGCGCTTACTGCGCTGGCGGACAGCGCGCGCGCGCGCGCGCG 205
|||||
252 GluSerLysGlyArgAlaHisProGlnValAspProLysLeuLeuAspLy 268
|||||
104 GAGTCCAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155
|||||
268 sleuHisGluThrPheHisGluProAsnLysLysPhePheLysLeuValG 285
|||||
54 ACTGACGAATATTTTCATGAGGCAATATGAGAAATTTCTGAGCTGTG 5
|||||
285 ly 285
||
4 GC 3

```

seq_name: /SID5/qrqdata/geneseq/geneseq-emb1/NA2001A.DAT:AA148186

seq_documentation_block:

ID AA148186 standard; DNA; 924 bp.

XX AA148186;

XX 17-OCT-2001 (first entry)

XX Probe #16872 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180412.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GH-0024283.

XX (MOLE+) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPT; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID NO 16872; 654bp; English.

XX

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX
SQ Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 9 Other;

alignment_scores:

Quality: 1335.50 Length: 284

Ratio: 4.928 Gaps: 1

Percent Similarity: 95.423 Percent Identity: 87.324

alignment_block:

US-09-557-262-2_COPY_21_307 x AA148186/rev ..

Align seg 1/1 to reverse of AA148186 from 1 to 924

```

2 ProAlaAlaProGlyProGlyLeuLysGlnGlnGluLeuArgLysVa 18
|||
851 CTTTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805
|||||
18 llelleleuProGluAspThrGlyGluGlyThrAlaSerAsnGlySer 35
|||
804 GGGGAGCGCTCCAGGATGAGTGGGTGATGGGTGAGGAGGAGGAG 755
|||||
35 hrGlnGlnLeuProGlnThrIlellelleGlyValArgLysGlyThr 51
|||||
754 CCCAGCAGTCCCGCAGACCATCATCCGCGTCCCGCAGCGCGCAGC 705
|||||
52 ArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaG 68
|||||
704 CCGGCACTCTTGGAGATGTTAGGCTGACAGGAGAGATGGAGGAG 655
|||||
68 uasnGluValHisPhePheAspTrpGluGluHisTyrSerClnGly 85
|||||
654 GAAAGAGGTCATTTCTTCGATGGGAGAGAGATTTAGGCACTGG 605
|||||
85 lyTrpTrpLeuThrGlnMetProPheSerSerProHisGlnLeuThr 101
|||||
604 GCTGGTACCTCAGCCAGATGGCTTCTCTCTGSCCACCACACAGCT 555
|||||
102 GluLysThrProAlaTyrPheThrSerProLysValProGluArg 118
|||||
554 GAGAGACCCCGCGGTATTTACGCTCGCCCAAAGTGGCTGAGCGAG 505
|||||
118 sSerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProSer 135
|||||
504 CAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 455
|||||
135 luArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGln 151
|||||
454 AGCGGCTGCTATCTGCTACACCCAAAGTGTCTACAAACACATGCA 405
|||||
152 HisLysProTyrProPheGluAspLeuLeuMetArgAspGlyArg 168
|||||
404 CAAAGAGGCTACCGGTATGAGAGAGAGAGAGAGAGAGAGAGAG 355
|||||
168 uasnleuasptryltsalaleuasnargserleuTyHisAlaHisMet 185
|||||
354 CAATGTGCAATACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
|||||
185 euasntrpleuargphepheleuGlyHisIleHisIleValAspGly 201
|||||
304 AGAACTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 255
|||||
202 AspArgLeuLeuArgAspProPheProGluLeuLysValGluArgP 218
|||||
254 GAGTCCAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205
|||||
218 cLeuLysLeuSerProGlnLysAlaSerAsnPhelThrAsnLys 235

```



```

PD 09-AUG-2001.
XX
XX 30-JAN-2001: 2001WO-US00569.
XX
XX 04-FEB-2000: 2000US-0180312.
XX
XX 26-MAY-2000: 2000US-0207456.
XX
XX 30-JUN-2000: 2000US-0608408.
XX
XX 03-AUG-2000: 2000US-0632366.
XX
XX 21-SEP-2000: 2000US-0234687.
XX
XX 27-SEP-2000: 2000US-0236359.
XX
XX 04-OCT-2000: 2000US-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR.
XX
XX WPI: 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 1: SEQ ID NO 3682: 638pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;
XX
alignment_scores:
XX      Quality: 748.00      Length: 150
XX      Ratio: 5.123      Gaps: 0
XX
XX Percent Similarity: 97.333      Percent Identity: 89.333
XX
alignment_block:
XX US-09-557-262-2_COPY_21_307 x ABA55377/rev ..
XX
XX Align seq 1/1 to reverse of: ABA55377 from: 1 to: 452
XX
XX 69 AsnGluValHisPheAspTrpGluHisTyrSerGlnGlyLeuI 85
XX
XX 451 AACGAGGTCCACTTCTTCCAGTGGGAGGACATTACAGCCACGCTTGGG 402
XX
XX 85 YTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 102
XX
XX 401 CTGGTAACTGAGCTAAATGAGTCTCTCTGATCAACAGCTACAGTGG 452
XX
XX 102 LuLysThrProAlaTyrPheThrSerProLysValProGluAArgHis 118
XX
XX 351 AGAAGACCCGCGCTATTTCATATGAGCAAAATGCTTGAGAGAGTATA 402
XX
XX 119 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuArgAspProSerG 135
XX
XX 301 ACATCAACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 252
XX
XX 135 uArqValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 152
XX
XX 251 GAGATGATATGAGTAAATCAATATGATGATATCAATATGATGATG 202
XX
XX 152 LysProTyrProProLysGluAspLeuLeuMetArgAspGlyArgLeu 168
XX
XX 201 ACAAGCCCTACCGCTCATGAGAGTCTGCTGCTGCTGCTGCTGCTGCT 152
XX
XX 169 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
XX
XX 151 AATGTGGACTACAGGCGCTCAAGCGGAGGCTTACCAAGCTGTCACATGCA 102

```

```

185 uAsnTrpLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 202
XX
XX 101 GAACTGATGATGTTTTCCTCCGCTGGGACATCAATATGATGATG 52
XX
XX 202 sPArgLeuLeuIleArgAspProPheProGluLeuGlnLysValGluArgPhe 218
XX
XX 51 ACCGCTCATCATGAGGAGCCCTCTCCCTGAGATCCAAAAGGTCGAGAGTTTC 2
XX
seq_name /SIN55/q3data/qgeneset/genesetp-emb1/NA2001A.DAT ABA25103
XX
seq_documentation_block:
XX ID ABA25103 standard; DNA; 452 BP.
XX
XX AC ABA25103;
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #3569 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
XX Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001: 2001WO-US00666.
XX
XX 04-FEB-2000: 2000US-0180312
XX
XX 26-MAY-2000: 2000US-0207456.
XX
XX 30-JUN-2000: 2000US-0608408.
XX
XX 03-AUG-2000: 2000US-0632366.
XX
XX 21-SEP-2000: 2000US-0234687.
XX
XX 27-SEP-2000: 2000US-0236359.
XX
XX 04-OCT-2000: 2000US-0024263
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR.
XX
XX WPI: 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 1: SEQ ID NO 3569; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;
XX
alignment_scores:
XX      Quality: 748.00      Length: 150
XX      Ratio: 5.123      Gaps: 0
XX
XX Percent Similarity: 97.333      Percent Identity: 89.333
XX
alignment_block:

```

US-09-557-262-2_copy_21_307 x AKA25103/rev

Align seq 1/1 to reverse of: AKA25103 from: 1 to: 452

```

69 AsnGluValHisPheAspTrpGluHisTyrSerGlnGlyLeuG 85
|||||
451 AAGCAGGTCATCTCTGAGATGGGAGAGATATACAGCCACGGCTGGG 402

85 yTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 102
|||||
401 CTGTGTATCTTATGACATATGCTTCTGATGATGATGATGATGATGATG 452

102 LuLysThrProAlaTyrPheThrSerProLysValProGluArgHis 118
|||||
351 AGAAAGCCGCTGATTTTATGATGATGATGATGATGATGATGATGATG 402

119 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 135
|||||
301 AGCATCAACCGCTGACATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 352

135 uArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 152
|||||
251 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 202

152 isLysProTyrProIleGluAspLeuLeuMetArgAspGlyArgLeu 168
|||||
201 ACAAAGCTACCTGCTGATGATGATGATGATGATGATGATGATGATGATG 152

169 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
|||||
151 AATGTGACTATGAAGGCTGCTGACCGTATGCTGCTGCTGCTGCTGCTGCTG 102

185 uAsnTrpLeuArgPheProGluGlyHisHisIleValAspGlyA 202
|||||
161 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 152

202 spArgLeuIleArgAspProIleGluAspLeuLeuLeuLeuLeuLeuLeu 218
|||||
51 ACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2

```

seq_name: /SIDS5/acdata/geneseq/geneseq-emb1/NA2201A.DAT.AAK03618

seq_documentation_block:

ID AAK03618 standard; DNA; 452 BP.

AAK03618;

05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 3609.

Human; brain expressed exon: gene expression analysis; probe:
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer; ss.

Homo sapiens.

W0200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001W0-0500667.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0608408.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0234687.

04-OCT-2000; 2000US-0234687.

(MOLFE-) MOLECULAR DYNAMICS INC.

PI Penn. SQ, Hanzel DK, Chen W, Rank DK;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
PI brains -

XX Example 4; SEQ ID NO: 3609; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:

Quality: 748.00 Length: 150
Ratio: 5.123 Gaps: 0
Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_copy_21_307 x AAK03618/rev

Align seq 1/1 to reverse of: AAK03618 from: 1 to: 452

```

69 AsnGluValHisPheAspTrpGluHisTyrSerGlnGlyLeuG 85
|||||
451 AAGCAGGTCATCTCTGAGATGGGAGAGATATACAGCCACGGCTGGG 402

85 yTrpTyrLeuThrGlnMetProPheSerProHisGlnLeuThrValG 102
|||||
401 CTGTGTATCTTATGACATATGCTTCTGATGATGATGATGATGATGATG 452

102 LuLysThrProAlaTyrPheThrSerProLysValProGluArgHis 118
|||||
351 AGAAAGCCGCTGATTTTATGATGATGATGATGATGATGATGATGATG 402

119 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 135
|||||
301 AGCATCAACCGCTGACATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 352

135 uArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 152
|||||
251 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 202

152 isLysProTyrProIleGluAspLeuLeuMetArgAspGlyArgLeu 168
|||||
201 ACAAAGCTACCTGCTGATGATGATGATGATGATGATGATGATGATGATG 152

169 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
|||||
151 AATGTGACTATGAAGGCTGCTGACCGTATGCTGCTGCTGCTGCTGCTGCTG 102

185 uAsnTrpLeuArgPheProGluGlyHisHisIleValAspGlyA 202
|||||
101 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 152

169 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
|||||
151 AATGTGACTATGAAGGCTGCTGACCGTATGCTGCTGCTGCTGCTGCTGCTG 102

185 uAsnTrpLeuArgPheProGluGlyHisHisIleValAspGlyA 202
|||||
101 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 152

202 spArgLeuIleArgAspProPheProGluLeuLeuLeuLeuLeuLeuLeu 218
|||||
51 ACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2

```

seq_name: /SIDS5/acdata/geneseq/geneseq-emb1/NA2201A.DAT.AAK03618

seq_documentation_block:

ID AAK29079 standard; DNA; 452 BP.

AAK29079;

05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 3609.

Human; brain expressed exon: gene expression analysis; probe:
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer; ss.

Homo sapiens.

W0200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001W0-0500667.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0608408.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0234687.

04-OCT-2000; 2000US-0234687.

(MOLFE-) MOLECULAR DYNAMICS INC.

US-09-557-262-2_COPY_21_307 x AA145025/rev ..

Align seq 1/1 to reverse of. AA135025 from. 1 to: 452

```
69 AsnGluValHisPhePheAspTrpGluGluHisTyrSerGlnClyLeuG1 85
|||||
451 AACGAGGTCGACCTTCGAGATGAGAGCATTAAGAGGCTTGGG 402
85 yTrpPtyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 102
|||||
401 CTGGTAACTAGCCAGATGAGGCTTCGCTGAGGACACAGCCACAGCTGG 352
102 LuLysThrProAlaTyrPheThrSerProLysValProGluArgIleHis 118
|||||
351 AGAAGACCCCGCGGTATTTTCACGTCGCGCAAAAGTGGCTGAGCGAGTCTAC 302
119 SerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProSerG1 135
|||||
301 AACATGAAACCGGTCATCCGGCTGCTGCTCATCTCGGAGACCCGTCGGA 252
135 uArgValIleuSerAspTyrThrGluValIleuTyrAsnHisLeuGlnIysH 152
|||||
251 CCGCGTCTCTATCTGACCTACACCCCAAGCTGTCTACAAACCACATCCAGAGC 202
152 isLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeu 168
|||||
201 ACAAGCCCTACCCGTCGATCGAGGAGTCTCTGTCGCGCATGCCAGGCTC 152
169 AsnLeuAspTyrIysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
|||||
151 AATGTGATTAAGAGCTCTCAAGCTGTAATCTATACAGTGCACATGCA 102
185 uAsnTrpLeuArgPhePheProLeuGlyHisIleHisIleValAspGlyA 202
|||||
101 GAATGGCTGCGCTTTTCGCGGTCGCGGACATCCACATTGTGGACGGCG 52
202 spArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 218
|||||
51 ACCGCTCATCAGGAGCCCGCTTCCCTCAGATCCAAAAGGTGAGAGGTTTC 2
```

OM 01: US-09-557-262-2_copy_21_307.p2n.rge to GenEmbl * out_format . pls

Date: Jul 17, 2002 5:47 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODE=framer.p2n model -DRV=xlp
-O/-cgn2 125SP10_Spool2050557_21_307.p2n.rge -GAPOP=12.000
-DH=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINANCH=0.100 -LAP=10.000 -LAPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XAP=10.000 -XAPEXT=0.500
-XAP=6.000 -XAPEXT=7.000 -XAP=10.000 -XAPEXT=0.500
-DELOP=6.000 -DELEX=7.000 -START=1 -MATRIX=blomsm62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=20 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pls
-NOPM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=unlimited
-USER-US09557262_CGNL1_11342_NFE01-6_ICPU-3_LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT_THREADS=1
```

Search information block:

```
Query: US-09-557-262_2_COPY_21_307
Query length: 287
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 7102.980000
```

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_ro:BC009133	+ 1544.00	2684.05	3.9e-141	1657	HC009133 Mus musculus, heparan
gb_ro:AF019385	+ 1544.00	2684.08	3.9e-141	1685	AF019385 Mus musculus heparan
gb_hlg:AC084416.3	+ 1544.00	2641.03	9.6e-139	110000	Continuation (4 of 6) of AC0
gb_hlg:AC098713	+ 1544.00	2633.92	2.4e-138	320128	AF098713 Mus musculus chromo
gb_hlg:AF177430	+ 1514.00	2637.35	1.5e-138	9466	AF177430 Rattus norvegicus chr
gb_pr:AF019386	+ 1345.50	2338.74	6.7e-122	1305	AF019386 Homo sapiens heparan
gb_pr:AC073991	+ 1345.50	2289.43	3.7e-119	159838	AC073991 Homo sapiens HAC ch
gb_pr:AC006240	+ 1338.50	2278.72	1.9e-118	165998	AC006240 Homo sapiens chromo
gb_pat:AX3131978	+ 745.50	1282.56	4.4e-63	2115	AX3131978 Sequence 18 from Pat
gb_pr:AL335498	+ 745.50	1237.89	1.4e-60	165549	AL335498 Human DNA sequenc
gb_om:CF3388534	+ 685.00	1191.01	5.6e-58	519	AF086544 Canis familiaris mRNA
gb_pr:AF105374	+ 592.00	1014.41	4.3e-48	1908	AF105374 Homo sapiens heparan
gb_pr:AF105375	+ 592.00	1014.41	4.3e-48	1968	AF105375 Homo sapiens heparan
gb_ro:AF105376	+ 582.00	994.60	2.9e-47	1666	AF105376 Mus musculus b. glycos
gb_pr:AF105377	+ 582.00	994.60	2.9e-47	1666	AF105377 Homo sapiens heparan
gb_pr:AF105378	+ 582.00	994.60	2.9e-47	1666	AF105378 Homo sapiens heparan
gb_pr:AC023723	+ 563.00	959.96	3.7e-45	2546	AF105376 Homo sapiens heparan
gb_pr:AC023723	+ 563.00	959.96	3.7e-45	2546	AF105376 Homo sapiens heparan
gb_hlg:AF105378	+ 540.00	922.12	5.3e-43	32311	AF105378 Homo sapiens heparan
gb_hlg:AC017405	+ 450.50	726.26	1.3e-32	93745	AC017405 Prosopithecus m. m. m.
gb_in:AC008337	+ 450.50	720.81	8.7e-32	16480	AC008337 Prosopithecus m. m. m.
gb_in:AE003511	+ 450.50	713.57	2.2e-31	44807	AE003511 Prosopithecus m. m. m.
gb_hlg:AC027402	+ 437.00	674.23	2.7e-29	16872	AC027402 Homo sapiens chromo
gb_pr:AC006287	+ 435.00	675.05	3.1e-29	188636	AC006287 Homo sapiens chromo
gb_hlg:AC098715	+ 425.00	672.19	4.4e-29	249269	AC098715 Mus musculus chromo
gb_hlg:AC093359	+ 424.00	671.79	4.7e-29	218502	AC093359 Mus musculus chromo
gb_pr:AC093511	+ 423.00	671.51	4.9e-29	184244	AC093511 Homo sapiens chromo
gb_hlg:AL603889	+ 412.00	646.85	1.1e-27	320066	AL603889 Mus musculus chromo
gb_hlg:AC021445	+ 410.50	648.86	8.9e-28	204621	AC021445 Mus musculus chromo
gb_in:GELH15803	+ 409.50	674.00	2.7e-29	12172	AF098996 Canis familiaris chro
gb_pr:AF098997	+ 404.50	668.69	7.1e-29	24810	AF098997 Canis familiaris chro
gb_pr:AF098998	+ 404.50	668.69	7.1e-29	24810	AF098998 Homo sapiens heparan
gb_pr:HS43946	+ 402.00	650.34	7.3e-28	41255	AF017723 Human DNA sequenc
gb_pr:AC005363	+ 399.00	658.94	5.2e-27	75108	AF005464 Homo sapiens chromo
gb_hlg:AC012140	+ 399.00	626.53	1.6e-26	251947	AF012140 Homo sapiens chromo
gb_hlg:AC005475	+ 394.50	624.04	2.1e-26	149060	AF005475 Homo sapiens chromo
gb_pr:AC005224	+ 392.00	618.50	4.3e-26	166647	AC005224 Homo sapiens chromo
gb_in:AF175689	+ 348.00	577.48	8.4e-24	4947	AF175689 Prosopithecus m. m. m.
gb_pr:AF0156429	+ 344.50	573.57	1.4e-23	3359	AF0156429 Homo sapiens NDST4 mR
gb_ro:AB036434	+ 337.00	565.17	4.1e-23	2732	AF036838 Mus musculus NDST4 mR
gb_pr:AF076605	+ 336.50	562.54	5.7e-23	2478	AF076605 Homo sapiens heparan

```
gb_pr:AF074524 + 330.50 561.84 6.1e-23 2188 AF074524 Homo sapiens hepar
gb_in:ELF5461 + 330.50 536.42 1.6e-21 98054 ELF5461 Oryzomys latipes eleq
gb_hlg:HS076601 + 330.50 551.22 3.1e-22 3221 HS076601 Homo sapiens heparan
gb_hlg:G07044 + 330.50 551.22 3.4e-22 3024 G07044 SRR023167 Human Hom
gb_hlg:AF221095 + 327.50 547.47 3.3e-22 2783 AF221095 Mus musculus hepar
```

seq_name: qb_ro:BC009133

seq_documentation_block:

Accession: BC009133

DEFINITION: Mus musculus, heparan sulfate (glucosamine) 3'-O-sulfotransferase 1,

clone W61-11450 IMAGE:415049, mRNA, complete cds.

ACCESSION: BC009133

VERSION: 1

KEYWORDS: house mouse,

SOURCE: house mouse,

ORGANISM: Mus musculus,

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Scurionanthi; Muridae; Murinae; Mus.

REFERENCE: 1 (bases 1 to 1657)

TITLE: Direct Submission

JOURNAL: Strausberg, R.

COMMENT: Submitted (05-JUN-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,

USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgabs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I M A G E Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: cdna@bcm.tmc.edu

Villation, D.K., Luna, P.A., Hale, S.M., Hulyk, S.L., Xie, X., Garcia,

A.M., Hellmeyer, M., Telford, H., Redfearn, A., Hock, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I M A G E Consortium/LLNL at: <http://image.llnl.gov>

Series: IPAK Plate: 16 Row: 0 Column: 17

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA at 2518970.

Location/Qualifiers

1..1657

Source

Organism: Mus musculus

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

Accession: BC009133

```

267  ptyrLeuHisGluTyrPheHisSerProAsnLysLysPhePheLysLeuV 284
|||||
1144  TAAATCGCAGCAATCTTCATCAGCCAAATAGAAATTTTTCAGGCTCG 1194
|||||
284  alcGlyAsnThr 287
|||||
1194  TGCGGAGAGACA 1204

seq_name: gb_ro:AF019385
seq_documentation_block:
LOCUS      AF019385
DEFINITION Mus musculus heparan sulfate D-glucosaminyl 3-O-sulfate precursor (cstII) mRNA, alternatively spliced, complete
ACCESSION AF019385
VERSION    AF019385.1  GI:2618970
KEYWORDS   .
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Fukuiyota; Metazoa; Chordata; Craniata; Vertebrata;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
1 (bases 1 to 1685)
LOCART,D. and Rosenberq,R.D.
TITLE       Molecular cloning and expression of mouse and human e
heparan sulfate D-glucosaminyl 3-O sulfate transferase
J. Biol. Chem. 272 (44), 28008-28019 (1997)
98010647
2 (bases 1 to 1685)
Shworak,N.W., Liu,J., Fritze,L.M.S., Schwartz,J.J.,
LOCART,D. and Rosenberq,R.D.
Direct Submission
Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street
02139, USA
FEATURES             Location/Qualifiers
     source           1..1685
                     /organism="Mus musculus"
                     /strain="C3H/An"
                     /db_xref="taxon:10090"
                     /cell_line="clonal 1 cell line LTA clone 1A"
                     Shworak et. al., 1996, J. Biol. Chem. 271:2
                     /note="poly(A)+ RNA isolated from postconfu
                     1..322
                     /gene="30st1"
                     /note="alternatively spliced; this version i
                     Class 1 mRNA"
                     1..1685
                     /gene="30st1"
                     216..217
                     /gene="30st1"
                     /note="splice variant site; many additional
                     (class 2 to class 9) contain extra 5' GTC
                     these nucleotides (extra 86 to 1629 bp)"
                     323..382
                     /gene="30st1"
                     323..1258
                     /gene="30st1"
                     /function="rate limiting enzyme for the syn
                     antieoquant heparan sulfate"
                     /note="heparan sulfate sulfotransferase; t
                     interluminal Golgi resident protein (retens
                     unknown)"
                     /codon_start=1
                     /product="heparan sulfate D-glucosaminyl
                     3-O-sulfotransferase-1 precursor"
                     /protein_id="AA884.467.1"
                     /db_xref="GI-2618971"
                     /translation="MTLLGAGVLLVAVQPOLVHSPAAPGCHLH
                     /TGCTASNGTSSQHPOTLTITGVKKGGRALLEMLSLHSDVAAAT
                     GLOWYITOMPFSFPHQITGVKTYAYFTSPKVDPERHSMNTITRQ
                     YVQVYVNHLAGKHRYPPYEDTLTLMHPLNLLYKALNSLYKALNSLYHAIH
                     VQGDPLTRPPEPTQKVEPFTKTSQVINA-SNEYENKTKGVCVLT

```



```

/product="3-O-sulfotransferase"
/protein_id="AAC09283.1"
/db_xref="gi:957244"
/translation="MTLLLLGAVLLVAQPOLVPSIAPAGGLKQQLLRKVIILPED
TGEAAATNGSTGQDPDTITTRVKGCTRALLEMUSLIPDVAANEVIFTFWEPIYSQ
GLGWYLTOMPSFSPHQLTVEKTPAYFTSPKVPKTHSMNPTIRLLLIIRDPSEVLSQ
YTOVLYNHQKHKPYPPIDLIJMRGRLNVDYKALNKSLSYAHMMIWLKFPPLVCHH
VQDREIRDPPELQKVERFKLSPIINASNPFYKTKGFCILRUSCKDKCLHRSKCK
AUPQVDPKLLDLKHEVEREPNKKFKLVGRFDWI"
BASE COUNT      224  292  224  196
ORIGIN

```

```

alignment_scores:
  Quality: 1514.00      Length: 287
  Ratio: 5.294          Gaps: 0
  Percent Similarity: 99.652  Percent Identity: 97.909

```

```
alignment_block:
US-09-557-262-2_Copy_21_307 x AF177430

```

```
Align seq 1/1 to: AF177430 from: 1 to: 936
```

```

1  HISPProAlaAlaProGlyProGlyLeuGlyGlnGlnGlnLeuLeuArgly 17
|||||
61  CACCGCGCTGCTCTGCTGCGCGCGGACCAACACGACGCGCTTCTGAGGAA 110
|||||
17  sValIleIleLeuProGluAspThrGlyGlyGlyThrAlaSerAsnGly 34
|||||
111  GTCGATTATTCGCCGAGACACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 160
|||||
34  crThrGlnGlnLeuProGlnThrIleIleIleGlyValArglyGly 50
|||||
161  GCAACCAAGAGAGTGCACACAAATATATATATTGATTTGATTTGATTTG 210
|||||
51  ThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAla 67
|||||
211  ACCGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 260
|||||
67  aGluAsnGlnValHisPhePheAspThrGluGluHisTyrSerGlnGly 84
|||||
261  TCAAAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310
|||||
84  cGlyIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 100
|||||
311  TGGATGTGATCTACCCACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
|||||
101  ValGlyIleThrProAlaTyrPheThrSerProIleValProGluArg 117
|||||
361  CTACAGACAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 410
|||||
117  eHisSerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeu 134
|||||
411  CACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 460
|||||
134  crGluArgValIleuSerAspIleThrGlnValIleuTyrAsnHisIleu 150
|||||
461  CGGAGAGTGTGATATGATGATGATGATGATGATGATGATGATGATGAT 510
|||||
151  LysHisIleProIleProIleGluAspIleuMetArgAspGlyArg 167
|||||
511  AAGCACAAACCTTACCCACCCATTCAGCAGCAGCAGCAGCAGCAGCAG 560
|||||
167  gLeuAsnLeuAspTyrIleValAlaLeuAsnArgSerLeuTyrHisAla 184
|||||
561  GCTCAACGTCGATACAGGCTCTCAACCCAGCCCTGATCCATGCGACACA 610
|||||
184  crLeuAsnThrLeuArgPhePheProLeuGlyHisIleHisIleValAsp 200
|||||
611  TGTGTAATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
|||||
201  GlyAspArgLeuLeuArgAspThrPheGluThrGlnGlyValGluArg 217
|||||

```

```

661  GGGGACCGGTTTCAACAGAGACCGCTTCCCTCAGATTCAGAAAGTGGAAAG 710
|||||
217  oPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnL 234
|||||
711  GPPGCTTAAAGTGTCTCAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 760
|||||
234  ySerTrpSerPheTyrCysLeuArgAspSerGlyGlyAspAspAspAsp 250
|||||
761  AAACCAAGGGTTCATGCTGCTGGGAGCAGTGGCAAGAGACAGCTGCTTA 810
|||||
251  HisLeuSerLysGlyArgAlaHisProGlnValArgProLysLeuLeuAs 267
|||||
811  CATGAGTCCAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
|||||
267  rLysLeuHisGlnTyrPheHisGluProAsnLysLysPhePheLysLeu 284
|||||
861  TAACTGCAAGAAATATTTTCTGAGAGCAATATAGAAATTTTTCAGGCTAG 910
|||||
284  aGlyArgThr 287
|||||
911  TGGGACAGAAVA 921

seq_name: qb_pr:AF019386
seq_documentation_block:
LOCUS      AF019386                1305 bp    mRNA    linear    191 14 Nov 1997
DEFINITION Homo sapiens heparan sulfate 3-O-sulfotransferase-1 precursor
ACCESSION  AF019386
VERSION    AF019386.1    GI:2618972
KEYWORDS   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 1305)
AUTHORS   Shworak, N.W., Liu, J., Fritze, D.M.S., Schwartz, J.J., Zhang, L.,
            Logeart, D. and Rosenberg, R.D.
            Molecular cloning and expression of mouse and human cDNAs encoding
            heparan sulfate D-glucosaminyl 3-O sulfotransferase
            J. Biol. Chem. 272 (44): 28008-28019 (1997)
            98010647
REFERENCE  2 (bases 1 to 1305)
AUTHORS   Shworak, N.W., Liu, J., Fritze, D.M.S., Schwartz, J.J., Zhang, L.,
            Logeart, D. and Rosenberg, R.D.
            Direct Submission
            Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
            02139, USA
FEATURES             Location/Qualifiers
     source            1..1305
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Isolated from lamda Triplex Brain cDNA library
                     from Clontech"
     gene              1..1305
                     /gene="3OST1"
     5'UTR            1..118
                     /gene="3OST1"
                     /note="3OST1"
     seq_peptide       119..178
                     /gene="3OST1"
                     /note="3OST1"
     CDS               119..1042
                     /gene="3OST1"
                     /function="rate limiting enzyme for synthesis of
                     antileucinant heparan"
                     /note="heparan sulfate sulfotransferase; 3-OST-1;
                     interluminal Golgi resident protein (retension mechanism
                     unknown)"
                     /codon_start=1
                     /product="heparan sulfate 3-O-sulfotransferase-1
                     precursor"
                     /protein_id="AAC04388.1"
                     /db_xref="gi:2618973"
                     /translation="MAALLLCAGVLLVAQPOLVPSIAPAGGLKQQLLRKVIILPEDWIRG

```


TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (08-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 159838)
Waterston,R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 159838)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 26, 2001 this sequence version replaced gi.8480614
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/qsc/>
Contact: sapient@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0512120

NOTE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:
The RP11-11 human HAC library was made from the blood of one male donor, as described by Otagawa, K., Wood, P.Y., Zhao, B., Fritzsche, E., Tateo, M., Calanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-669F3; the clone sequenced to the right is AC066230. Actual start of this clone is at base position 1 of RP11-512120; actual end is at base position 159838 of RP11-512120.

FEATURES
source
Location/Qualifiers
1..159838
organism="Homo sapiens"
db_xref="taxon:9606"
chromosome="4"
map="4"
clone="RP11-512120"
clone_lib="RP11-11"
2..775
repeat_region
266..317
repeat_region
776..792
repeat_region
/rpt_family="MERL_type"
/rpt_family="A-rich"
/rpt_family="AT-rich"

repeat_region
845..2469
/rpt_family="L1"
repeat_region
2470..2766
/rpt_family="Alu"
repeat_region
2769..3291
/rpt_family="MERL_type"
repeat_region
3321..3441
/rpt_family="L1"
repeat_region
3628..3655
/rpt_family="AT-rich"
repeat_region
4297..4409
/rpt_family="MIR"
repeat_region
4815..5120
/rpt_family="Alu"
repeat_region
5094..5114
/rpt_family="(A)n"
repeat_region
5203..5546
/rpt_family="L2"
repeat_region
8632..8739
/rpt_family="GA-rich"
repeat_region
9549..9890
/rpt_family="ERV1"
repeat_region
10107..10349
/rpt_family="MIR"
repeat_region
11853..12152
/rpt_family="MALR"
repeat_region
12209..12371
/rpt_family="L1"
repeat_region
12305..12390
/rpt_family="L2"
repeat_region
12605..12876
/rpt_family="L1"
repeat_region
12877..13185
/rpt_family="Alu"
misc_feature
13162..13185
/note="similar to Homo sapiens EST BF183502 (NM:0106181)"
repeat_region
13186..13221
/rpt_family="L1"
repeat_region
13224..13488
/rpt_family="L1"
repeat_region
13489..13750
/rpt_family="Alu"
repeat_region
13751..14291
/rpt_family="L1"
repeat_region
14332..14380
/rpt_family="MERL_type"
repeat_region
15690..15714
/rpt_family="(T)n"
repeat_region
15873..16172
/rpt_family="Alu"
repeat_region
16173..16300
/rpt_family="Alu"
repeat_region
16910..16942
/rpt_family="AT-rich"
repeat_region
17130..17499
/rpt_family="ERV1"
repeat_region
17303..17479
/rpt_family="A-rich"
repeat_region
17553..17621
/rpt_family="(CATATA)n"
repeat_region
17920..17970
/rpt_family="GA-rich"
repeat_region
18178..18658
/rpt_family="L1"
repeat_region
18644..18953
/rpt_family="L1"
repeat_region
19026..19155
/rpt_family="L2"
repeat_region
19061..19084
/rpt_family="AT-rich"
repeat_region
19167..19441
/rpt_family="L1"

```

repeat_region 19638..19749
/rpt_family="L1"
repeat_region 19850..20600
/rpt_family="L1"
repeat_region 20251..20494
/rpt_family="MER2_type"
repeat_region 21233..22285
/rpt_family="L1"
repeat_region 23307..23479
/rpt_family="MER1_type"
repeat_region 24526..24779
/rpt_family="TCTA)n"
repeat_region 24570..24849
/rpt_family="Alu"
repeat_region 25680..25700
/rpt_family="(TA)n"
repeat_region 26134..26688
/rpt_family="ERV1"
repeat_region 26944..27375
/rpt_family="MALR"
repeat_region 27885..28356
/rpt_family="L2"
repeat_region 28532..28641
/rpt_family="L2"
repeat_region 28723..28942
/rpt_family="MIR"
repeat_region 29079..29542
/rpt_family="L2"
repeat_region 29566..29888
/rpt_family="L2"
repeat_region 29899..30411
/rpt_family="ERV1"
repeat_region 30439..30539
/rpt_family="L2"

alignment_scores:
Quality: 1345.50 length: 296
Ratio: 4.929 Gaps: 1
Percent Similarity: 95.455 Percent Identity: 87.413

alignment_block:
US-09-557-262-2_copy_21_307 x AC073991.rev ..
Align seq 1/1 to reverse of: AC073991 from: 1 to: 159838

2 ProAlaLaProGlyProGlyLeuLysGlnGlnGlnLeuArqLysVa 18
||||| ||||| ||| ||||| ||||| ||||| |||||
155036 CCGCCGCGCGCGCGGAG CTAAGCCAGCCAGGTTCTGCGCAAGC 154940

18 lilelleLeuProGluAspThrGlyGluGlyThrAlaSerAsnGlySerf 35
||||| ||||| ||||| ||||| ||||| ||||| |||||
154989 GAGGAGGCTGTAAGATAGGAGGATGAGGAGGAGGAGGAGGAGGAG 154940

35 hrgGlnGlnProGlnThrIlelleGlyValArgLysGlyGlyThr 51
||||| ||||| ||||| ||||| ||||| ||||| |||||
154939 CCGACAGTGGCGCAGACCAATCATCATCGCGGCGCGCAGCGCGCAG 154890

52 ArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaGl 68
||||| ||||| ||||| ||||| ||||| ||||| |||||
154889 CCGCGACCTGCTGGAGATGCTGAGCTGAGCGGCGCGCGCGCGGGA 154840

68 uAsnGluValHisPheAspThrGluGluHisTyrSerGlnGlyLeuG 85
||||| ||||| ||||| ||||| ||||| ||||| |||||
154839 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154790

85 lylTyrTyrLeuThrGlnMetProPheSerSerProHisGlnGlnThrVal 101
||||| ||||| ||||| ||||| ||||| ||||| |||||
154789 GGTGCTATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154740

102 GluLysThrProAlaTyrPheThrSerProLysValProGluArqLleH 118
||||| ||||| ||||| ||||| ||||| ||||| |||||
154739 CAGAGACCGCGCGGATCTTCACGCGCGGCAAAAGGAGGAGGAGGAG 154690

```

```

118 sserMetAsnProThrIleArgLeuLeuLeuIleLeuArqAspProSerG 135
||||| ||||| ||||| ||||| ||||| ||||| |||||
154889 CAGATATATATATATATATATATATATATATATATATATATATAT 154840

135 lAlaValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151
||||| ||||| ||||| ||||| ||||| ||||| |||||
154639 AGCGCGTCTATCTGACTACACCAAGTGTCTACACCAACACATCCAG 154590

152 HisLysPheTyrProPheIleGluAspLeuLeuMetArqAspGly 168
||||| ||||| ||||| ||||| ||||| ||||| |||||
154589 CAGAAAGGATATATATATATATATATATATATATATATATATATAT 154540

168 uAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMet 185
||||| ||||| ||||| ||||| ||||| ||||| |||||
154539 CAATATGATATATATATATATATATATATATATATATATATATATAT 154490

185 euAsnTrpLeuArqPhePheProLeuGlyHisIleHisIleValAspGly 201
||||| ||||| ||||| ||||| ||||| ||||| |||||
154489 ACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154440

202 AspArqLeuIleArqAspProPheProGluIleGlnLysValGlnArqP 218
||||| ||||| ||||| ||||| ||||| ||||| |||||
154439 GAGCGGCTCATCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154390

218 eLeuLysLeuSerProGlnIleAspAlaSerAspPheTyrPheAspLys 235
||||| ||||| ||||| ||||| ||||| ||||| |||||
154389 CATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154340

235 hrLysGlyPheTyrCysLeuArqAspSerGlyLysAspArqCysLeuHis 251
||||| ||||| ||||| ||||| ||||| ||||| |||||
154339 CCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154290

252 GluSerLysGlyArgAlaHisProGlnValAspProLysLeuLeuAspLy 268
||||| ||||| ||||| ||||| ||||| ||||| |||||
154289 GAGTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154240

268 sLeuHisClnTyrThrHisGlnTrpAsnLysTyrPheLysLeuValG 295
||||| ||||| ||||| ||||| ||||| ||||| |||||
154239 ATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 154190

285 lYArqThr 287
|||||
154189 GCAGAAC 154182

seq_name: gb.pr.AC006230
seq_documentation_block:
LOCUS AC006230
DEFINITION Homo sapiens chromosome 4 clone C0287J14 map 4p16, complete
sequence.
AC006230.11 GI:4966389
VERSION AC006230.11
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166998)
AUTHORS Stone,N.E., Schmutz,J.J., Sheng,J., Pennacchio,L.A., Cox,D.P. and
Myers,R.M.
TITLE Direct Submission
JOURNAL Unpublished
PEPREFERENCE 2 (bases 1 to 166998)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.P. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-1998) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
PEPREFERENCE 3 (bases 1 to 166998)
AUTHORS Stone,N.E., Schmutz,J.J., Sheng,J., Pennacchio,L.A., Cox,D.P. and
Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Department of Genetics, Stanford Human
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA

```



```

repeat_region /note-"15 copies 6 mer tatata 72% conserved"
7185..7239
/note-"18 copies 2 mer ta 77% conserved"
8044..8192
/note-"MIR repeat: matches 22..237 of consensus"
8200..8415
/note-"MER63B repeat: matches 209..436 of consensus"
8416..8708
/note-"AluSq repeat: matches 1..293 of consensus"
8709..8954
/note-"MER63B repeat: matches 1..209 of consensus"
8983..9031
/note-"MIR repeat: matches 193..241 of consensus"
9130..9173
/note-"MIR repeat: matches 76..119 of consensus"
9884..10031
/note-"L1M33 repeat: matches 6740..6807 of consensus"
10112..10172
/note-"M11A1 repeat: matches 1..365 of consensus"
10538..10734
/note-"MER63A repeat: matches 22..224 of consensus"
10747..10856
/note-"L1M4 repeat: matches 3607..3737 of consensus"
10885..11176
/note-"L2 repeat: matches 2471..2607 of consensus"
11273..11438
/note-"AluLo repeat: matches 144..207 of consensus"
11439..11742
/note-"AluSc repeat: matches 1..304 of consensus"
11743..11788
/note-"AluLo repeat: matches 97..144 of consensus"
11885..11967
/note-"L2 repeat: matches 2637..2718 of consensus"
13808..13919
/note-"MIR repeat: matches 35..147 of consensus"
14580..14874
/note-"AluSc repeat: matches 1..295 of consensus"
15659..15980
/note-"match: STS: EM-60413"
15712..15779
/note-"34 copies 2 mer ca 79% conserved"
15716..15775
/note-"10 copies 6 mer cacaca 81% conserved"
17139..17258
/note-"FLAM-A repeat: matches 1..122 of consensus"
18171..18700
/note-"MLT1D repeat: matches 2..505 of consensus"
18863..18939
/note-"AluLo/FLAM repeat: matches 43..124 of consensus"
18940..19217
/note-"AluLo repeat: matches 1..304 of consensus"
19286..19447
/note-"MLT1H repeat: matches 118..264 of consensus"
19588..19694
/note-"MLT1H repeat: matches 345..483 of consensus"
20090..21304
/note-"Tiger3b repeat: matches 1..1230 of consensus"
22028..22075
/note-"L2 repeat: matches 2458..2508 of consensus"
23080..23346
/note-"AluSq repeat: matches 41..308 of consensus"
29225..29588
/note-"THEIC repeat: matches 1..371 of consensus"
29765..30063
/note-"AluSq repeat: matches 1..299 of consensus"
31420..31708
/note-"AluSq repeat: matches 1..292 of consensus"
32985..33283
/note-"AluSq repeat: matches 1..301 of consensus"
34330..34786
/note-"L2 repeat: matches 2248..2703 of consensus"
44858..45006
/note-"MIR repeat: matches 79..248 of consensus"

```

```

repeat_region 35193..35341
/note-"MIR repeat: matches 48..194 of consensus"
36032..36094
/note-"MER5A repeat: matches 88..152 of consensus"
36125..36410
/note-"AluSq repeat: matches 1..306 of consensus"
37642..37713
/note-"MIR repeat: matches 106..180 of consensus"
38972..39030
/note-"MIR repeat: matches 148..199 of consensus"
39074..39210
/note-"MIR repeat: matches 31..178 of consensus"
39269..39573
/note-"AluSq repeat: matches 1..303 of consensus"
39700..39823
/note-"L1B2 repeat: matches 5930..6152 of consensus"
40412..40941
/note-"L2 repeat: matches 1813..2407 of consensus"
40961..41060
/note-"MIR repeat: matches 131..233 of consensus"
41058..41472
/note-"L2 repeat: matches 2406..2710 of consensus"
42145..42381
/note-"MER63 repeat: matches 745..998 of consensus"
42430..44048
/note-"L1PA7 repeat: matches 4526..6144 of consensus"
44072..44167
/note-"MER63 repeat: matches 2..734 of consensus"
44787..44901
/note-"L2 repeat: matches 2609..2747 of consensus"
45261..45368
/note-"MER45 repeat: matches 73..1027 of consensus"
47949..48247
/note-"AluSq repeat: matches 2..308 of consensus"
48321..48466
/note-"MIR repeat: matches 37..191 of consensus"
48586..49514
/note-"TIGER1 repeat: matches 1489..2417 of consensus"
49515..49828
/note-"AluSp repeat: matches 1..508 of consensus"
49829..50088
/note-"TIGER1 repeat: matches 1230..1489 of consensus"
50148..50693
/note-"L1PA16 repeat: matches 5603..6157 of consensus"
50716..51171
/note-"TIGER1 repeat: matches 4..475 of consensus"
51605..51816
/note-"L1M88 repeat: matches 5956..6171 of consensus"
52358..52403
/note-"MER67A repeat: matches 449..543 of consensus"
52411..52488
/note-"L1TR10A repeat: matches 13..98 of consensus"

alignment_scores:
Quality: 745.50 Length: 290
Ratio: 3.404 Gaps: 3
Percent similarity: 75.517 Percent identity: 48.966

alignment_block:
US-09-557-262-2_COPY_21_307 x AL355498/rev
Align seq 1/1 to reverse of: AL355498 from: 1 to: 165549
1 HisProAlaAlaProGlyProGlyLeuLysGlnIleG61uLeuLeuAraTy 17
|||||
27845 CACGAGTTCGGGAAGGCAAGCGTTCACAGGAGGACGATTCGCTTCACCA 27796
17 sValIleIleLeuLeuG61uAspThrGlyGlyThrAlaSerAsnGlyS 34
|||||
27795 CCGTGGTTC..... 27789
34 G61uThrGlnIleLeuProGlnThrIleIleIleGlyValArgLysGlyGly 50
|||||

```

```

27788 ... CAGAGTCTTAAAGGATATATATCAATGAGATAGAAAGAGAG 27744
51 ThrArgAlaIatLeuLeuMetLeuSerLeuHisProAspValAlaAlaI 67
|||||
27743 ACAAGGGCCCTGCTTGAATATCAACATACATCCGACAGIACAAGC 27694
|||||
67 aduAsnGluValHisPhePheAspTrpGluGluHisTyrSerGlnGly 84
|||||
27693 CTCTCAAGAAATCCACTTTTGTGATATGATCAACAATATGCTAAGCC 27644
|||||
84 euGlyTrpTrpLeuThrGlnMetProPheSerProHisGlnLeuThr 160
|||||
27643 TTGATGTGTATAGAAAAAGATGTTTTTTTATGCTTACAAATTA 27544
|||||
101 ValGluTyrThrProIatIyrPheThrSerProTyrValProGluArg 117
|||||
27593 ATTGAAAGAGCCAGCATATTTTATCACAGAGGAGCTTCCAGAAAG 27544
|||||
117 eHisSerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAspPro 134
|||||
27543 TTACAAATCAACTCATCAACTTCTTGATCAITCTCAGGAGCCAA 27494
|||||
134 euGluArgValLeuSerAspTrpThrGlnValLeuTyrAsnHisLeuGln 150
|||||
27493 CCACAAAGATATTTTATTAATATTAATATTAATATTAATATTA 27444
|||||
151 lysHisTyrProTyrProTolIleGluAspLeuLeuMetArgAspGly 166
|||||
27443 AAGAACAATAATTTTACAAAGTTTCAGAAAGCTGCCATCAAGC 27394
|||||
167 ... ArqLeuAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHis 182
|||||
27393 ATGCCAAGTCAACACAAATCAAGAGAGTAGAAGACCAAGCATCAC 27344
|||||
182 eHisMetLeuAsnTrpPheArgPheProLeuGlyHisIleHisIle 198
|||||
27343 AACATCTAGAGAGATTTTAAATATTTTAAATATTAATATTAAT 27294
|||||
199 ValAspGlyAspArgLeuIleArgAspProPheProGluIleGlnLys 215
|||||
27293 CTCCATGCGACATCCGCTCATACAGGAGCTCTGACAGATCAGT 27244
|||||
215 lGluArgPheLeuLysLeuSerProGluIleAsnAlaSerAsnPro 232
|||||
27243 GCAGAGTTCCTTAATATTCGCTCCACAGCATTAAGTCAATATTA 27194
|||||
232 heAsnLysThrLysGlyPheCysLeuArg...AspSerGlyLysAsp 247
|||||
27193 TCAATCTTAAAGAGATTTTATTAATTAATTAATTAATTAATTA 27144
|||||
248 ArgCysLeuHisGluSerLysGlyArgAlaHisProGlnValAspPro 264
|||||
27143 AAGTGGCTGGGCGGAGCAAGAGGAGGATTCATCCAGAGTGGACCC 27094
|||||
264 sLeuLeuAspLysLeuHisGluTyrPheHisGluProAsnLysLys 281
|||||
27093 TGTCATTAATAATCCCAATCTTTCATCCCTTTTAATCAAAAAAT 27044
|||||
281 heLysLeuValGlyArgThr 287
|||||
27043 ACCAGATCACTGGAGAGACA 27024

```

seq_name: gb_om:CFA388534

seq_documentation_block:

LOCUS CFA388534

DEFINITION Canis familiaris mRNA for partial heparan sulfate

3-O-sulfotransferase-1 precursor (3OST1 gene).

ACCESSION AJ388534

VERSION 3OST1.1 GI:5441572

KEYWORDS 3OST1 gene; heparan sulfate 3-O-sulfotransferase-1 precursor.

SOURCE dog.

ORGANISM Canis familiaris

```

Eukaryotic Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 519)
Eichler, H., Merican, B., Pouillon, V., Christophe Robertus, C. and
Christophe, D.
A method for the large-scale cloning of nuclear proteins and
nuclear targeting sequences on a functional basis
Anal. Biochem. 264 (2), 231-235 (2000)
2 (bases 1 to 519)
Christophe, D.
Direct Submission
Submitted (23-JUN-1999) Christophe D., I.R.I.B.H.N., U.L.B., route
de Lennik 808, Bldg. C, 1070 Bruxelles, BELGIUM
Related sequence: AF019386.
FEATURES
Source
1..519
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone="Bcl"
/tissue_type="thyroid"
1..519
/gene="3OST1"
<!-->519
/feature="3OST1"
/codon_start=3
/product="heparan sulfate 3-O-sulfotransferase-1
precursor"
/protein_id="CAH48853.1"
/db_xref="GI:5441572"
/translation="AVLLAAQLPLVPSDFPAAGAEPSAAATLQGEGRGGAAGNSAQQ
LPGTLLIGVKGGI KALLPMISHPDAVAANFVFFIWEHYSGGLQWYLGOMPPSS
PHQIVKTKPAYFTSPKVPVHSMNPGLRIILLIRDSERVLSUYIQVFNHVKKK
PAPSIRFETVRD"
BASE COUNT 81 a 199 c 169 g 70 t
ORIGIN

```

```

alignment_scores:
Quality 48.00 Length 519
Ratio: 4.692 Gaps: 2
Percent Similarity: 87.952 Percent Identity: 80.120
alignment_block:
US-09-557-262-2_copy_21_307 x CFA388534
Align seq 1/1 to: CFA388534 from: 1 to: 519
2 ProAlaAlaProGly.....ProGlyLeuLysGlnGlnGlnLeuLeuAr 16
|||||
45 CACCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 71
16 nLysValIleIleLeuProGluAspThrGlyGlyThrAlaSerAsnG 33
|||||
72 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
33 ySerThrGlnGlnLeuProGlnThrIleIleGlyValArgLysGly 49
|||||
121 GCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170
50 GlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAla 66
|||||
171 GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 220
66 aAlaGluAsnGluValHisPhePheAspTrpGluGluHisTyrSerGln 83
|||||
221 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 270
83 yLeuGlyCysTrpTrpTrpLeuThrGlnMetProPheSerProHisGln 99
|||||
271 GACTGCGCTGGTACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 320

```


[illegible]

seq_name: qb_ro:AF168992

seq_documentation_block:

AF168992	1656 bp	mRNA	linear	POD 31-OCT-1999
DEFINITION	Mus musculus D-glycosaminyl 3-O sulfotransferase-3b mRNA, complete cds.			
ACCESSION	AF168992			
VERSION	AF168992.1	GI:6164709		
KEYWORDS	house mouse			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 1656)			
AUTHORS	Shukla, D., Liu, J., Hlaiklock, P., Shworak, N.W., Hall, X., Esko, J.D., Cohen, G.H., Eisenberg, R.J., Rosenberg, R.D. and Spear, P.G.			
TITLE	A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry			
TEMPNAL	Cell 99 (1), 13-22 (1999)			
MELLINE	9444300			
PURMED	10520990			
REFERENCE	2 (bases 1 to 1656)			
AUTHORS	Shukla, D., Liu, J., Hlaiklock, P., Shworak, N.W., Rosenberg, R.D. and Spear, P.G.			
TITLE	Direct Submission			
TEMPNAL	Submitted (12-MIL-1999) Microbiology/Immunology, Northwestern University, 633 E. Chicago Avenue, Chicago, IL 60611, USA			
FEATURES	Location/Qualifiers			
SOURCE	1..1656			
	/organism="Mus musculus"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10099"			
	356..1528			
CDS	/note="3-OST-3b"			
	/codon_start=1			
	/product="D-glycosaminyl 3-O-sulfotransferase-3b"			
	/protein_id="AAAF04505.1"			
	/db_xref="GI:6164710"			
	/translation="MGQPISSGSGSCIVDPCEPLPGPPPPPPVPRKLLALLFAMLCIWL YMFILYSCAGSCIAAKGILLIGSSKALHAQPLIVLAFNPKSKPKPPFAPPAVNSLAACK DKTVAGSQSRQSEAPDSVSPISPPSCAGSKQLQALILCVKKGQGTALPLPTIRVH ETKPAVNAPEHFFPTSGYGLAWDPMPTLKGITWKPSTYFTPEAPARISMS KILVLVPPPTVTRATISGYTILSPEDPLFESLIFPNSALIDTSNSATGGLY AKHLEMLRHFPLGVLVSGSEPLVSDPAEGLFPQVDEGLKPTITTHKFFYFQTKF PCLKAGSKGCPHCIAKTKGKAHPTIAREVILPQIROPYFPFNPKFYOMGRDPCWD"			
BASE COUNT	331 a	556 c	493 g	286 t
ORIGIN				

```
alignment_scores:
  Quality: 582.00
  Pairs: 2.969
  Gaps: 4
  Percent Similarity: 74.809
  Percent Identity: 44.656
```

alignment block:

attachment_block: US-09-557-262-2 COPY 21 307 x AF168932

Align seq 1/1 to: AF168992 from: 1 to: 1666

```
    } SerAsnGlySerTArgInGlnIeuProGlnThrIleIleIeIvAlAr 47
```

[illegible]

47 σ | vsG | vG | vtrArTA | a | en | enG | uMet | enSor | enHisproAspV 64

794 :AAGGGGCGTGAATCTCAATTGTCGTAAACCTTAACCTTAACTC

44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Figure 1 is a schematic representation of the experimental design. It shows a flow from 'Study 1' to 'Study 2'. Study 1 involves 'Pretest' and 'Main Study'. Study 2 involves 'Pretest' and 'Main Study'. The 'Main Study' in Study 2 is divided into 'Control' and 'Intervention' groups. The 'Intervention' group is further divided into 'Intervention 1' and 'Intervention 2'.

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2001, 08:46:39 ; Search time 106.58 Seconds
(without alignments)
2317.987 Million cell updates/sec

Title: US-09-557-262-3

Perfect score: 1305

Sequence: 1 cgcgcctcagtaattgaagg... gaactaaatatttactga 1305

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size: 16

Total number of hits satisfying chosen parameters: 149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA: *
1: /cgn2.6/prodata/2/ina/5A_COMB.seq: *
2: /cgc2.6/prodata/2/ina/5A_COMB.seq: *
3: /cgn2.6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2.6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2.6/prodata/2/ina/PTUS_COMB.seq: *
6: /cgn2.6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	19	1.5	558	4	US-08-998-416-906	Sequence 906, Appl
2	18	1.4	1451	4	US-09-276-531-28	Sequence 28, Appl
3	18	1.4	7812	4	US-09-368-590-1	Sequence 1, Appl
4	18	1.4	10747	2	US-08-147-777-1	Sequence 1, Appl
5	18	1.4	10747	3	US-08-452-872-1	Sequence 1, Appl
6	18	1.4	10747	5	PCT-US93-03985-1	Sequence 1, Appl
7	18	1.4	24979	2	US-08-147-777-3	Sequence 3, Appl
8	18	1.4	24979	3	US-08-452-872-3	Sequence 3, Appl
9	18	1.4	24979	5	PCT-US93-03985-3	Sequence 3, Appl
10	17	1.3	36	1	US-08-311-760A-96	Sequence 96, Appl
11	17	1.3	36	1	US-08-311-760A-271	Sequence 271, Appl
12	17	1.3	36	1	US-08-311-760A-272	Sequence 272, Appl
13	17	1.3	36	1	US-08-311-760A-273	Sequence 273, Appl
14	17	1.3	36	2	US-08-774-310-96	Sequence 96, Appl
15	17	1.3	36	2	US-08-774-310-271	Sequence 271, Appl
16	17	1.3	36	2	US-08-774-310-272	Sequence 272, Appl
17	17	1.3	36	2	US-08-774-310-273	Sequence 273, Appl
18	17	1.3	49	1	US-07-885-689A-26	Sequence 26, Appl
19	17	1.3	51	1	US-07-885-689A-13	Sequence 13, Appl
20	17	1.3	387	1	US-08-458-084-8	Sequence 8, Appl
21	17	1.3	387	1	US-08-205-508-8	Sequence 8, Appl
22	17	1.3	387	2	US-08-482-148-5	Sequence 5, Appl
23	17	1.3	387	5	PCT-US95-02944-5	Sequence 5, Appl
24	17	1.3	387	5	PCT-US95-02945-8	Sequence 8, Appl
25	17	1.3	576	1	US-07-885-689A-28	Sequence 28, Appl
26	17	1.3	576	1	US-08-093-383-4	Sequence 4, Appl
27	17	1.3	579	1	US-07-885-689A-36	Sequence 36, Appl

Sequence 37, Appl
Patent No. 5514646
Sequence 6, Appl
Sequence 3, Appl
Sequence 7, Appl
Sequence 12, Appl
Sequence 880, Appl
Sequence 55, Appl
Sequence 57, Appl
Sequence 24, Appl
Sequence 25, Appl
Sequence 10, Appl
Sequence 58, Appl
Sequence 88, Appl
Sequence 139, Appl
Sequence 85, Appl

579 1 US-07-885-689A-47
600 6 5614646-37
601 1 US-07-764-655D-6
601 1 US-07-801-164A-3
604 1 US-07-764-655D-7
675 4 US-08-209-603E-12
680 4 US-09-196-293-12
735 4 US-08-998-416-880
821 4 US-08-235-836C-55
821 4 US-08-235-836C-56
821 4 US-08-235-836C-57
822 3 US-08-541-807A-24
822 3 US-08-542-807A-25
822 4 US-08-235-836C-10
822 4 US-08-235-836C-58
822 4 US-08-235-836C-88
822 4 US-08-235-836C-129
825 4 US-08-235-836C-85

ALIGNMENTS

RESULT 1
US-08-998-416-906
Sequence 906, Application Us/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippseu, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebschunq, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER US/08/998.416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/BOOK NUMBER: PF/5-10306/A/00-1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 906:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1565UP



US-08-998-416-906

Query Match 1.5%; Score 19; DB 4; Length 558;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 tttaaaagcatttaagg 1130
Db 295 TTTAAAAAGCATTTAAGG 313

RESULT 2

US-09-276-531-28/C
Sequence 28, Application US/0942/16531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lynn E. Murty, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1451 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HNT2GT01
CLONE: 491493
US-09-276-531-28

Query Match 1.4%; Score 18; DB 4; Length 1451;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 915 tacatgagtcctaaagccc 932
TACATGAGTCTCAAGGCC 279

RESULT 3
US-09-368-590-1
Sequence 1, Application US/09368590
Patent No. 6187563
GENERAL INFORMATION:
APPLICANT: Solimena, Michele
TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES
FILE REFERENCE: 101918-200 (OCR-941)
CURRENT APPLICATION NUMBER: US/09/368,590
CURRENT FILING DATE: 1999-08-04
EARLIER APPLICATION NUMBER: 60/395,657
EARLIER FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 7812
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(5879)
NAME/KEY: unsure
LOCATION: (100) (102)
NAME/KEY: unsure
LOCATION: (1021)...(1023)
NAME/KEY: unsure
LOCATION: (2266)...(2268)
US-09-368-590-1

Query Match 1.4%; Score 18; DB 4; Length 7812;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 gcgcggtgctgctggtgg 155
Db 2045 gcgcggtgctgctggtgg 2062

RESULT 4

US-08-147-777-1
Sequence 1, Application US/08147777
Patent No. 5914265
GENERAL INFORMATION:
APPLICANT: Koop, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KSPATIN V1 EXPRESSION VECTORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:


```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:475268"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/xex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpd total tetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not 1. Site.2: Eco RI; 1st strand cDNA
was primed with a Not 1 - oligo(dT) primer [5',
TGTTCACCAACTCGAGGAGGAGCGCCGGGAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpd and 2
14.5dpd embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not 1 and
cloned into the Not 1 and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
136 a 134 c 114 g 124 t

```

Query Match 30.1%; Score 508; DB 1; Length 508;
 Best Local Similarity 100.0%; Pred. No. 4.8e-257;
 Matches 508; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	875	gacggctggcctgaaacctggactcaaaagctctcaaacccagcctgtacacatcacacatg	934
Db	1	GACGGTCGGCTGAACCTGGACTACAAAGGCTCTCAACCGCAGCCTGTACATGCACACATG	60
QY	935	ctgaaactggctggcgtttttcccgcttggggccacatccacattgtgtagcgaccgcctc	994
Db	61	CTGAACCTGGCTGGCTTTTTCCTCGTGGCGACATCCACATTGTGGATGGCAGCCGCTC	120
QY	995	atcagagaccctttcccttgagatccagaaggtcgaaagattcctgaagctttctccacag	1054
Db	121	ATCAGAGACCCCTTTCCCTGAGATCCAGAAGTCCGAAGATTCTGAGGTTTCTCCACAG	180
QY	1055	atcaacgcctcgaactctactttaacaaacaaagggtcttactgcctgcggacagt	1114
Db	181	ATCAACGGCTTGAACTTCTATTTAACAAACCAAGGGTTTCTACTGCTGCGGACAGT	240
QY	1115	ggcaaggaccgctgcttacacagagtcacaaagccggcgaccccaggtggatcccaaa	1174
Db	241	GGCAAGGACCCGCTGCTTACACGAGTCTCAAGAGCGCGGCGCACCCGAGTCCCAAA	300
QY	1175	ctacttgataactgcacgaatactttcatagacccaataagaaatttttcaagctcgtg	1234
Db	301	CTACTTTGATAACTTGCACGAATACTTTTCATGACGCAAAATGAAGAAATTTTCAAGCTCGTG	360
QY	1235	ggcagaacattcgactggcactgatttgcgctctcctagctcgaggacttttctgttgt	1294
Db	361	GGCAGAACTTCGACTGGCACTGATTTGATTTGGCGCTCGCTGAGCTCGGACITTCCTGTGTG	420
QY	1295	taactcttgggtgacatctaaaggggggagaaaataatttttaaagggcatttaagcta	1354
Db	421	TAACCTCTGGTGTACATCTGAAGGGGGAGGAAAAATTAATTTTAAAAAGGCATTTAAGCTA	480
QY	1355	taattatttgtlaaaaccccaaatgac	1382
Db	481	TAATTATTATTGATAAACCCCAAAATGAC	508

RESULT	4
WG2484	
LOCUS	WG2484 490 bp mRNA EST
DEFINITION	md72cl2.l1 Soares mouse embryo NDMEJ3.5.14.5 Mus musculus cDNA clone IMAGE:373942 5' similar to PIR:A49733 A49733
ACCESSION	glycosaminoglycan N-acetylglucosaminyl N-deacetylase ; , mRNA sequence. WG2484

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the State in the Development of the Economy	Journal of Economic Surveys	See page 100
2. The Impact of Globalization on the Labor Market	Journal of International Economics	See page 100
3. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics	See page 100
4. The Role of the State in the Development of the Environment	Journal of Environmental Economics	See page 100
5. The Impact of Globalization on the Environment	Journal of International Economics	See page 100
6. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics	See page 100
7. The Role of the State in the Development of the Environment	Journal of Environmental Economics	See page 100
8. The Impact of Globalization on the Environment	Journal of International Economics	See page 100
9. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics	See page 100
10. The Role of the State in the Development of the Environment	Journal of Environmental Economics	See page 100

FEATURES SOURCE

```

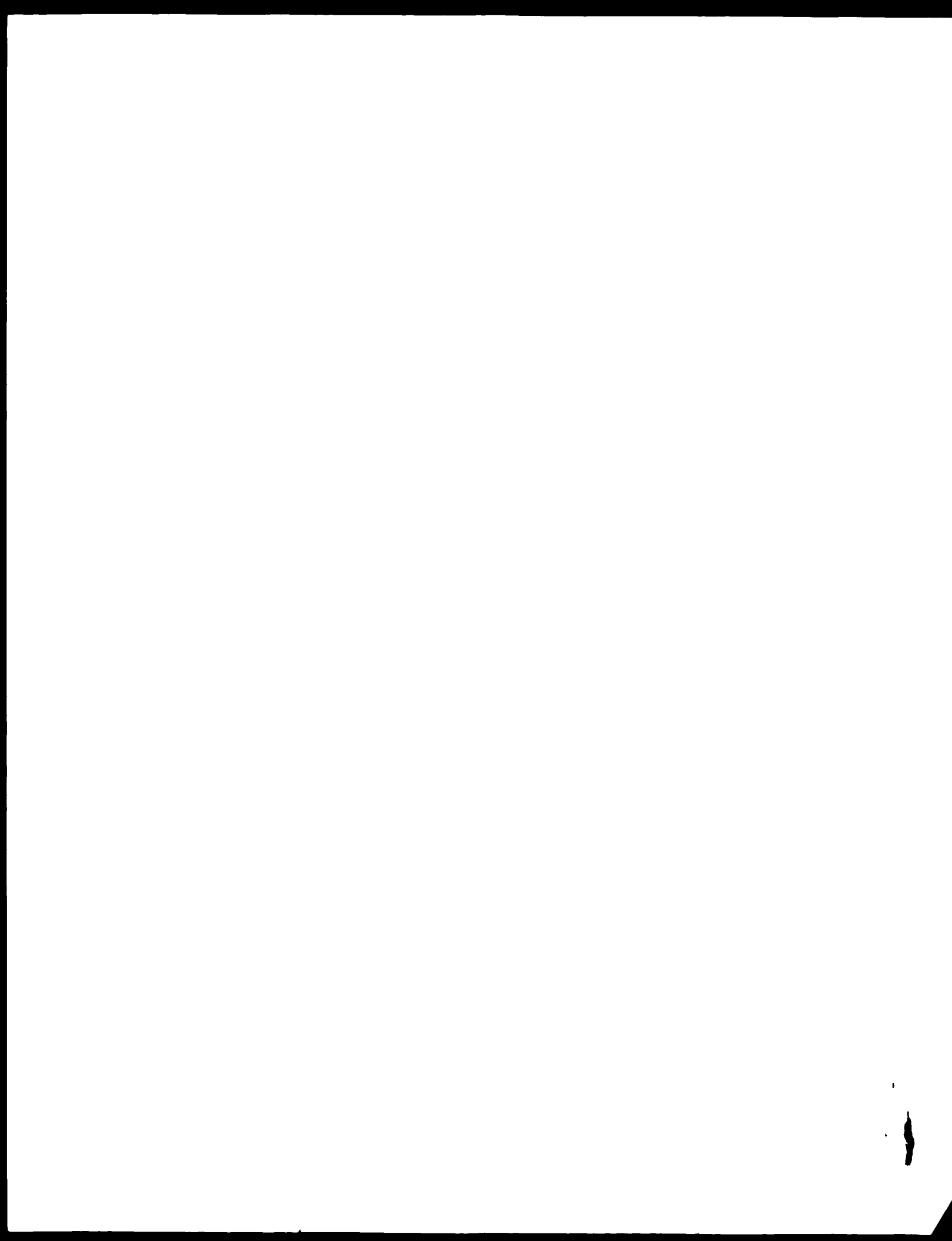
Location/Qualifiers
1. .490
/organism="Mus musculus"
/strain="c57Bl/6J"
/db_xref="taxon:10690"
/clone_image="IMAGE:373942"
/clone_lib="Soares mouse embryo NDME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total tetus"
/lab_host="DH10B"
/notes="Vector: pT73p-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI. 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCACTGAGTGGAGCGGCGCGGAATTTTCTTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minsu Koo, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Nvi I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fátima Bonaide. "
127 a 129 c 122 t

```

BASE COUNT
ORIGIN

Query Match:	
Best Local	5
Matches	490

Matches	490:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
QY	875	gacggtcgagcggaactgagctactaagaaggctctcaacgcgcgcgcgtgacactgacacatg	934						
DB	1	GACGCTCGGCTGAACTGGACCTGACTAAGAGGCTCTCAACCGCAGCGCTTGACATGACACATG	69						
QY	935	ctgaaactggctgcggttttttcccgatgagcgacatccacatttgagagcagcagccctc	994						
DB	61	CTGAACCTGGCTGGCTTTTTTCCCGTTGGCCACATCCAACTTGAGGATGAGGAGGCGCT	126						
QY	995	atcagagacaccccttcctcctgagatccaggaaggtcgaaagaattccctgaagctctctcaaga	1054						
DB	121	ATCAGAGCGCTTTCCTCGAATCCAGAAAGCTGGAAGATTCCGTAAGATTCCTCAGAG	189						
QY	1055	atcaacgcctctgaactctcaactttaaagaaacacagggctctcactgctgaggagacat	1114						
DB	181	ATCAAGCGCTCGAACTTCCTACTTTTAAACAAACCAAGGGCTTCTACTGCTCTGGGAGAGT	240						



361	TCACAGTATTAGATTACAGTTCGCCATATATAGTAGTATATATTTTCTACTGTTAAAT	420
QY		
1446	ggaggcgctttgtatgttttctatggtgtttaaacttggtatgatactctataatat	1505
Db		
421	GGAGGCGCTTTGTATTGTTTTTCATGGTGTGTTAAACATTGTG- -TATGCTCTATAATAT	478
QY		
1506	gaagaacttaactatgctgactgaataataagagatttttttctctggagaactcttt	1565
Db		
479	GAAGGAACCTTAACATATGCGACTGAAAAAATAAGAGATTTTTTCTCGAGACGTCTTT	538
QY		
1566	ttttgtgtgtgtttttaaataataataacctgctccaatcccaataagctctttggt	1625
Db		
539	TT-----TTTTTAATATTAATTAACCTGCCTCCCAATTCAAAATAGCTCTCTTGTT	586
QY		
1626	ttca--cctcctgtcaaatctataatctttttctgtctaaataattctatgggtatta	1681
Db		
587	TGCAACCCNCTTGCAAACTATAATCTTTTTTCTGCTTAAAAAAATTTATTGGTTTA	644

RESULT	9
AA041885	
LOCUS	508 bp mRNA
DEFINITION	mj05g11.1 Soares mouse embryo NMEL3.5 14.5 Mus musculus CDNA clone IMAGE:475268 5', similar to PIR:A49733 A49733 glycosaminoglycan N-acetylglucosaminyl N-deacetylase ;, mRNA sequence.
	03-SEP-1996

ACCESSION	AA041885
VERSION	AA041885.1
KEYWORDS	GI:1520040
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata, Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	1 (bases 1 to 508) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucabara, J., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and Waterston, R.
TITLE	The WashU-HHMI Mouse EST Project

unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:286012
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop. 490.
Location/Qualifiers
1..508
FEATURES
source

```

1. .308
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_name="IMAGE:475268"
/clone_lib="Soares mouse embryo NDME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpd total fetus"
/lab_host="DH10B"
/notes="(Vector: pT7T30-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5,
TGTACCAACATCGAGAGGGAGCGCCCGGAAATTTTTTTTTTTTTTTT
1-3]), on equal amounts of mRNA from 2 13.5dpd and 2
14.5dpd embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2] double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the NotI and Eco RI sites of the modified

```

	pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fátima Bonaldo.	"		
BASE COUNT	136 a	134 c	114 g	124 t
ORIGIN				

```

Query Match          30.1%; Score 508; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.e-122,
Matches 508; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

875	gacggtcggtggaacctgactacaaggctctcaaccgcagcctgtacctgcacaatg	934
QY		
1	GACGTCGGGTGAACCTGACTACAAAGCTCTCAACACGACGCTGATACATGACACATG	60
DB		
935	ctgaacctggctggctgtttttccggttgggtcgagatcgaattgtgagatgagctgc	994
QY		
61	CTGAACTCGGCTGGGTTTTTTCGCGTTGGGTCACATCTCCACATTTGTGTATGTCGCGTC	120
DB		
995	atcagagacccttctccctgagatccagaaaggtcgaagaattccctgaaagcttctccacag	1054
QY		
121	ATCAGAGACCTTTTCCTGAGATTAAGAAATCTGAAAGATTCCTGAAGCTTCTCCACAG	180
DB		
1055	atcaacgctctcgaactctactttaacaaacccaagggtctctactccctcgagacaagt	1114
QY		
181	ATCAACGCCCTCGAACCTTCTACTTTTAAACAACCAAGAGGTTCTACTGTGCTGGGACGT	240
DB		
1115	ggcagagaccgctgctttacacgagtcacaaggccggggcgaccccccagctgagatcccaaa	1174
QY		
241	GGCAAGGACCGCTGCTTACACGAGTCCAAAGGCGCGGCGACACCCACAGTGCATCCAAA	300
DB		
1175	ctacttgataaactgcacaaatacttctcagagacaaataagaaattttcaactcgtg	1234
QY		
301	CTACTTTATTAACCTGTCATGAATACATTCAAGAGCCAAATATAGAAATTTTCAAGCTCGTG	360
DB		
1295	ggcggaaacattccgaactgggcactgatttgcgtctctctaggctcggagatttctctgttgt	1294
QY		
361	GGCAGAACATTGCATCTGGTACATTTATTTGTTGTTGTTATATATGCGGACTTTTCCTGTTGT	420
DB		
1295	taactctggtgtacatctgaaugggggaggaataatatttataaaagcatllaaagcta	1354
QY		
421	TAACTTCCTGGTGATCATCTGAAGGGGGAGGAAATAATTTTAAAGAGGCAATTTAAAGCTA	480
DB		
1355	taatttattgttaaacccccaasatgac	1382
QY		
481	TAATTTATTTGTAAACCCCAAAATGAC	508
DB		

```

RESULT 10
RF940710
LOCUS
DEFINITION
    RF940710 651 bp mRNA
    nae24b08.xl NCI-CGAP v018 Homo sapiens cDNA clone IMAGE:343599, 4'
    similar to TR014752 G14792 HEPATOCYTE CARCINOMA CELL LINE cDNA IMAGE:343599, 4'
    PRECURSOR. ; mRNA sequence.
ACCESSION
    RF940710
VERSION
    RF940710.1 GI:12358030
KEYWORDS
    EST.
SOURCE
    human.
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Crustacea; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1. (bases 1 to 651)
AUTHORS
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
    Unpublished (1997)
COMMENT
    Contact: Robert Strausberg, Ph.D.

```

email: cypapou@u.washington.edu
Tissue Procurement: Christopher A. Muskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prance, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

STIC-Biotech/ChemLib

From:
Sent:
To:
Subject:

Shukla, Ram
Monday, August 27, 2001 1:14 PM
STIC-Biotech/ChemLib
09/557,262

Please search for:

nucleic acid of Seq ID NO 1
nucleic acid encoding SEQ ID NO 2
nucleic acid of Seq ID NO 3
nucleic acid encoding SEQ ID NO 4
Nucleic acid encoding residues 21-52, 260-269, 250-276, 53-311 or 21-307 of SEQ ID NO 2
Nucleic acid encoding residues 21-48, 256-265, 246-272, 49-307 or 21-303 of SEQ ID NO 2
Oligosearch (16 mer) for SEQ ID NO 1 and 3

Thanks.

Patent Examiner
AU 1632
Crystal Mail 1, Rm 12E03
USPTO
Phone: (703) 305-1677
Fax: (703) 746-3103

10E12
m3

Edward Grant
Technical Info Specialist
STIC / Biotech
CMI 12C14 Tel: 305-9203

49702



=) d his

(FILE 'HOME', ENTERED AT 17:38:56 ON 26 OCT 2001)

FILE 'MEDLINE', ENTERED AT 17:39:16 ON 26 OCT 2001

81 S GST-C OR O-SULFOTRANSF

870 S GST OF C-SULFOTRANSF OR 3-OST

870 S GST OF C-SULFOTRANSF OR 3-OST

11901 S HOST CELL

73984 S WHOLEP

183624 S FRAGMENT

1 S 12 AND 13 AND 14 AND 15

1 S 3-OST

1 S 12 NOT EX-1996

54 S HEPARIN SULFATE PROTEOGLYCAN OR HEPARIN SULPHATE PROTEOGLYCAN

1 S 13 AND 113

80 S 1210-TRANSF AND 13

1 S 1210-TRANSF AND 13 AND 3-OST

30 S 3-OST-TRANSF AND 3-OST

1 S 1210-TRANSF

1 S 1210-TRANSF

1 S 1210-TRANSF

1 S 1210-TRANSF

1 S 1210-TRANSF

1 S 1210-TRANSF

FILE 'CAPSUS', ENTERED AT 17:54:16 ON 26 OCT 2001

17:54:16 ON 26 OCT 2001

39 S 113

144 S 113

39 S 113 NOT EX-1996

1 S DUP REM 124 123 DUPLICATES REMOVED

FILE 'USPATENT', ENTERED AT 17:57:41 ON 26 OCT 2001

11 S 113

1300 F HID

0 S (NUCLEIC ACID OF DNA OF RNA) S) 126

7 S (NUCLEIC ACID OF DNA OF RNA) AND 126

L Number	Hits	Search Text	DB	Time stamp
1	6	3-ost or o-sulfoltransferase	USPAT; US-PGPUB; EPO; JPO; DERWENT	2001/10/26 18 08
7	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same host same vector	USPAT; US-PGPUB; EPO; JPO; DERWENT	2001/10/26 18 09
19	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same vector	USPAT; US-PGPUB; EPO; JPO; DERWENT	2001/10/26 18 09
25	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same host	USPAT; US-PGPUB; EPO; JPO; DERWENT	2001/10/26 18 09
13	2	(3-ost or o-sulfoltransferase) same nucleic ADJ acid	USPAT; US-PGPUB; EPO; JPO; DERWENT	2001/10/26 18 10
31	2	transgen\$ and (3-ost or o-sulfoltransferase)	USPAT; US-PGPUB; EPO; JPO; DERWENT	2001/10/26 18 10



L	Number	Hits	Search Text	DB	Time stamp
7		2	(3-ost or o-sulfotransferase) same nucleic ADJ acid	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:51
13		154108	(3-ost or o-sulfotransferase) and human or mouse	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:50
19		154109	3-ost or o-sulfotransferase and human or mouse	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:51
25		6	transgen\$ and (3-ost or o-sulfotransferase)	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:51
		10	3-ost or o-sulfotransferase	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:51
		6	3-ost or o-sulfotransferase	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:49
		0	(3-ost or o-sulfotransferase) same nucleic ADJ acid same host same vector	USPAT; EPO; JPO; US-PGPUB;	2001/10/26 18:09
		0	(3-ost or o-sulfotransferase) same nucleic ADJ acid same vector	USPAT; EPO; JPO; US-PGPUB;	2001/10/26 18:09
		0	(3-ost or o-sulfotransferase) same nucleic ADJ acid same host	USPAT; EPO; JPO; US-PGPUB;	2001/10/26 18:09
		2	(3-ost or o-sulfotransferase) same nucleic ADJ acid	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:50
		2	transgen\$ and (3-ost or o-sulfotransferase)	USPAT; EPO; JPO; US-PGPUB;	2002/07/25 14:51
		22101	signal ADJ sequence	USPAT; EPO; JPO; US-PGPUB;	2001/10/30 14:41
		3	("5541095").PN.	USPAT; EPO; JPO; US-PGPUB;	2001/10/30 14:41
		2	("5817487").PN.	USPAT; EPO; JPO; US-PGPUB;	2001/10/30 14:42
		0	("11 same 17").PN.	USPAT; EPO; JPO; US-PGPUB;	2001/10/30 14:42
		DERWENT		USPAT; EPO; JPO; US-PGPUB;	2001/10/30 14:42



STIC-Biotech/Chemlib

From: Shukla, Ram
Sent: Monday, July 15, 2002 1:39 PM
To: STIC-Biotech/Chemlib
Subject: 09/557,262

Please search for:
NA encoding 21-307 of SEQ ID NO 2
NA encoding 53-311 of SEQ ID NO 2
NA encoding 250-276 of SEQ ID NO 2
NA encoding 21-303 of SEQ ID NO 4
NA encoding 49-307 of SEQ ID NO 4
NA encoding 246-272 of SEQ ID NO 4.

Thanks

Patent Examiner
AU 1632
Crystal Mail 1, Rm 12E03
Mailbox 12E12
USPTO
Phone: (703) 305-1677
Fax : (703) 746-3103

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

TYPE OF SEARCH:

NA Sequences: 6
AA Sequences:
Structures:
Bibliographic:
Litigation:
Full text:
Patent Family:
Other:

Searcher:
Phone:
Location:
Date Picked Up:
Date Completed:
Searcher Prep/Review:
Clerical:
Online time:

VENDOR/COST (where applic.)
STN:
DIALOG:
Questel/Orbit:
DRLink:
Lexis/Nexis:
Sequence Sys.:
WWW/Internet:
Other (specify):

70881

